Rec'd PCT/PTC 19 JAN 2005 #2

REC'D 05 SEP 2003
WIPO PCT

P1 1057288

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

August 28, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/397,440***
FILING DATE: ***July 19, 2002***
RELATED PCT APPLICATION NUMBER: *PCT/US03/22001*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

M. Sias

M. SIAS
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

7-22-02 603[illegible] .071902

07/19/02 jc511 U.S. PTO

A/PR
Box

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53 (c)

Docket Number	UTSG:255USP1	Type a plus sign (+) inside this box →	+

jc973 U.S. PTO 60/397444

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Barrett	Alan		Galveston, Texas
McArthur	Monica		Galveston, Texas

TITLE OF THE INVENTION (280 characters max)

METHODS AND COMPOSITIONS CONCERNING ALTERED YELLOW FEVER VIRUS STRAINS

CORRESPONDENCE ADDRESS

FULBRIGHT & JAWORSKI L.L.P.
600 Congress Avenue, Suite 2400
Austin, Texas 78701
USA

ENCLOSED APPLICATION PARTS (*check all that apply*)

☒ Specification — *Number of Pages* 67 and Sequence Listing - 31 pages — ☒ Small Entity Statement

☒ Drawing(s) — *Number of Sheets* 7 — ☒ Other (specify) Seq Stm. w/disk, hard copy of seq. listing, ck $80; postcard

METHOD OF PAYMENT (*check one*)

☒ A check or money order is enclosed to cover the Provisional filing fees.

☒ The Assistant Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number: 50-1212/10205941/UTSG 255USP1 should any fees be missing or deficient.

☒ Pursuant to 37 CFR 1.53(g) this provisional application is being filed without a filing fee. Please send the "Notice to File Missing Parts" form pursuant to 37 CFR 1.53(g).

PROVISIONAL FILING FEE AMOUNT $80

☒ Applicants are entitled to Small Entity Status Pursuant to 37 C F R § 1.27

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No

☐ Yes, the name of the U S Government agency and the Government contract number are ________

CERTIFICATE OF EXPRESS MAILING
NUMBER EL 794534941 US
DATE OF DEPOSIT July 19, 2002

Respectfully submitted,

Charles P. Landrum
Reg. No. 46,855
FULBRIGHT & JAWORSKI L.L.P.
600 Congress Avenue, Suite 2400
Austin, Texas 78701
512-536-5674

Date: July 19, 2002

32425
PATENT TRADEMARK OFFICE

25189780.1

6039 440.071902

CERTIFICATE OF EXPRESS MAIL
NUMBER EL 794534941 US
DATE OF DEPOSIT July 19, 2002

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:
Alan Barrett and Monica McArthur

Serial No.: UNKNOWN

Filed: CONCURRENTLY HEREWITH

For: METHODS AND COMPOSITIONS CONCERNING ALTERED YELLOW FEVER VIRUS STRAINS

Group Art Unit: UNKNOWN

Examiner: UNKNOWN

Atty. Dkt. No.: UTSG:255USP1

STATEMENT AS REQUIRED UNDER 37 C.F.R. § 1.821(f)

BOX SEQUENCE
Commissioner for Patents
Washington, D.C. 20231

Commissioner:

Submitted herewith is a computer readable form and a paper copy of the sequence listing of those sequences in the captioned patent application. The computer readable form of the sequence listing is the same as the paper copy of the sequence listing. The sequence information provided in the Specification is also the same as the sequence listing of the enclosed computer readable and paper forms of the sequence listing.

Respectfully submitted,

Charles P. Landrum

Charles P. Landrum
Reg. No. 46,855
Agent for Applicants

FULBRIGHT & JAWORSKI L.L.P.
600 Congress Avenue, Suite 2400
Austin, Texas 78701
512/474-5201

Date: July 19, 2002

PATENT
UTSG:255USP1

PROVISIONAL
APPLICATION FOR UNITED STATES LETTERS PATENT
for
METHODS AND COMPOSITIONS CONCERNING ALTERED YELLOW FEVER VIRUS STRAINS
by

Alan Barrett
and
Monica McArthur

CERTIFICATE OF EXPRESS MAIL
NUMBER EL 794534941 US
DATE OF DEPOSIT July 19, 2002

25176161.1

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates generally to the fields of molecular biology and virology. More particularly, it concerns nucleic acid compositions and methods for using such compositions to develop Yellow Fever vaccines.

2. Description of Related Art

The disease Yellow Fever, caused by a member of the *Flaviviridae* family designated Yellow Fever (YF) virus, is prevented by the use of live attenuated vaccine known as 17D. The 17D virus was developed by passage of Yellow Fever virus wild-type strain Asibi (isolated in Ghana in 1927) in chicken tissue. The 17D vaccine is manufactured by six producers worldwide who jointly manufacture approximately 100-150 million doses annually. The 17D vaccine has an excellent safety record with only 21 reports of expression of a neurovirulent phenotype. In recent years there have been three separate reports (in Brazil, Australia and USA) of 17D expression of a viscerotropic phenotype, with cases of apparent Yellow Fever-type disease, causing concern and is threatening the use of 17D vaccine.

Tesh *et al.* reported studies on three YF strains in hamsters (Tesh *et al.*, 2001). Two strains became viscerotropic only following intraperitoneal inoculation of virus and multiple liver-to-liver passages in hamsters. One strain, Jimenez, (isolated in Panama in 1974 from a human case) was unusual in that it caused viscerotropic disease in hamsters without adaptation by passage in hamsters and killed a proportion of animals.

There exists a clear need for vaccines that will stimulate an immune response in a subject, while reducing the potential for expression of a virulent phenotype. Thus, methods and compositions useful for the production and use of improved vaccines would be beneficial.

SUMMARY OF THE INVENTION

Compositions and methods of the present invention include provisions for the improvement of *flavivirus* vaccines so that the risk of disease is reduced or eliminated. In certain embodiments the *flavivirus* is a Yellow Fever virus. In other embodiments, a virus may be an altered 17D, 17D-204, 17DD, or other Yellow Fever vaccines. In various other embodiments the vaccine may be a chimeric vaccine, as described herein. Chimeric refers to a viral genome, viral polypeptide or viral particle that contains a discernable portion(s) of at least two viruses or virus strains, and may also include portions of non-viral nucleic acids and/or polypeptides.

In various embodiments an isolated nucleic acid encoding a Yellow Fever virus with a viral genome that may include at least one of the following alterations: a) an alteration in the nucleic acid sequence resulting in an envelope protein (described below) with a histidine at amino acid 27; b) an alteration in the nucleic acid sequence resulting in an envelope protein with a glycine at amino acid 28; c) an alteration in the nucleic acid sequence resulting in an envelope protein with an alanine at amino acid 155; d) an alteration in the nucleic acid sequence resulting in an envelope protein with an arginine at amino acid 323; e) an alteration in the nucleic acid sequence resulting in an envelope protein with an arginine at amino acid 331; f) an alteration in the nucleic acid sequence resulting in a NS2A protein (described below) with an alanine at amino acid 48; or g) an alteration in the nucleic acid sequence resulting in a NS4B protein (described below) with an isoleucine at amino acid 98. Each of the alterations may be used in combination with each and every other combination of the remaining alterations and/or other alteration in a 5' or 3' noncoding region (NCR) and/or a core (C), a PrM, an M, an envelope (E), a NS1, a NS2A, NS2B, NS3, NS4A, 2K, NS4B, NS5 protein(s) and combinations thereof, each of which is described below. A nucleic acid sequence representative of a hamster passage 7 Yellow Fever virus sequence is presented in SEQ ID NO:1. A polypeptide sequence representative of a hamster passage 7 Yellow Fever virus sequence is presented in SEQ ID NO:2. SEQ ID NO:3 is a portion of SEQ ID NO:1 that encodes an envelop protein and SEQ ID NO:4 sets forth a polypeptide that represents a processed envelop

protein. The location of all other protein may be determined by analysis of the genbank sequences described below.

The nucleic acids of the invention may be RNA or DNA. In some embodiments where the nucleic acid is DNA transcription will be oriented so that an infectious RNA will typically be transcribed from the DNA.

In various other embodiments, a nucleic acid encoding all or part of a viral genome may include at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more alterations. Alterations may be innocuous or render the virus more or less immunogenic, replication competent, virulent or alter other characteristics of the virus. In certain embodiments, the nucleic acid the polynucleotide has a nucleic acid sequence as set forth in SEQ ID NO:1.

In other embodiments a nucleic acid comprising 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9,000, 10,000 or more and values there between of contiguous nucleotides of SEQ ID NO:1.

In yet other embodiments, a vaccine composition may include a Yellow Fever virus with a viral genome that includes at least one of the following alterations and may include any combination thereof: a) an alteration in the nucleic acid sequence encoding amino acid 323 of an envelope protein, wherein the first alteration requires more than one nucleotide change to encode an arginine; b) an alteration in the nucleic acid sequence encoding amino acid 27 of an envelope protein, wherein the second alteration requires more than one nucleotide change to encode a histidine; c) an alteration in the nucleic acid sequence encoding amino acid 28 of an envelope protein, wherein the second alteration requires more than one nucleotide change to encode a glycine; d) an alteration in the nucleic acid sequence encoding amino acid 155 of an envelope protein, wherein the second alteration requires more than one nucleotide change to encode an alanine; e) an alteration in the nucleic acid sequence encoding amino acid 331 of an envelope protein, wherein the second alteration requires more than one nucleotide change to encode an arginine; f) an alteration in the nucleic acid sequence encoding amino acid 48 of an NS2A protein, wherein the second alteration requires more than one nucleotide change to

encode an alanine; or g) an alteration in the nucleic acid sequence encoding amino acid 98 of an NS4B protein, wherein the second alteration requires more than one nucleotide change to encode an isoleucine. The envelop protein is encoded by nucleotides 974 to 2452 of SEQ ID NO:1 and corresponds to amino acids 286 to 778 of SEQ ID NO:2.

The Yellow Fever virus viral genome may include at least two, three, four, five, six, or seven alterations in any combination. Typically, the vaccine composition is in a pharmaceutically acceptable formulation. Additionally, the vaccine composition may include the 17D virus, 17D-204 virus, 17DD virus, or other viral variants with any combination of alterations incorporated therein.

In still other embodiments, a method for producing an attenuated Yellow Fever virus including introducing into a Yellow Fever virus genome a missense mutation that would require two nucleotide changes to encode a supervirulence amino acid is contemplated. An attenuated virus refers to a virus that has been modified or treated to reduce or eliminate its ability to cause disease.

In various embodiments, methods for producing a Yellow Fever virus vaccine may include: a) identifying a mutation that results in a missense mutation in a first Yellow Fever viral genome that is associated with an increased virulence of the virus; b) modifying an attenuated Yellow Fever viral genome by mutation of a codon associated with the missense mutation resulting in a reduced probability of reversion to a virulent phenotype. In certain embodiments, the method may include a missense mutation results in an envelope protein having an arginine at amino acid position 323 (SEQ ID NO:2) and may also include any combination of other alterations in the viral genome. The method may include modifying the attenuated Yellow Fever virus by substituting a second codon that encodes for a conservative amino acid change.

In other embodiments, a method for identifying a compound active against a viral infection including, but not limited to: a) providing a virus expressed from a viral construct comprising a nucleic acid encoding a Yellow Fever virus comprising an envelope protein with an arginine at amino acid 323; b) contacting said virus with a candidate substance; and c) comparing the infectious ability of the virus in the presence of said candidate substance with the infectious ability of the virus in a similar system in the absence of said candidate substance is contemplated. The method may also include a

nucleic acid encoding a virus with an envelope protein including, but not limited to a histidine at amino acid 27, a glycine at amino acid 28, an alanine at amino acid 155, and/or an arginine at amino acid 331, as well as any other combination of alterations. In certain embodiments a nucleic acid sequence is that set forth in SEQ ID NO:1 or a polynucleotide sequence as set forth in SEQ ID NO:2, or other related flaviviral sequences.

In various embodiments, methods of vaccination including, but not limited to administering to a subject a Yellow Fever virus with a viral genome that includes at least one of the following alterations: a) an alteration in the nucleic acid sequence encoding amino acid 323 of an/the envelope protein, wherein the first alteration requires more than one nucleotide change to encode an arginine; b) an alteration in the nucleic acid sequence encoding amino acid 27 of an/the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a histidine; c) an alteration in the nucleic acid sequence encoding amino acid 28 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a glycine; d) an alteration in the nucleic acid sequence encoding amino acid 155 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode an alanine; e) an alteration in the nucleic acid sequence encoding amino acid 331 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode an arginine; f) an alteration in the nucleic acid sequence encoding amino acid 48 of the NS2A protein, wherein the second alteration requires more than one nucleotide change to encode an alanine; or g) an alteration in the nucleic acid sequence encoding amino acid 98 of the NS4B protein, wherein the second alteration requires more than one nucleotide change to encode an isoleucine, as well as compositions used in vaccination are contemplated. The viral genome may also include at least a combination of two, three, four, five, six, seven or more alterations. The vaccine composition is typically in a pharmaceutically acceptable formulation. The vaccine composition may include, but not limited to having a 17D virus, 17D-204 virus, 17DD virus, or other Yellow Fever viral variants, as well as other viral strains and species. Methods of vaccination may include administration of an effective amount of a vaccine composition such that an immune response to virus is induced in a subject. In various embodiments, vaccination and

vaccine compositions may include adjuvants and other excipients, as well as additional antigen(s) that may induce an immune response(es) to the same or other pathogen, foreign body, or organism.

Various embodiments of the invention may include, but are not limited to a) nucleic acid compositions comprising all or part of the nucleotide sequence of the hamster p7, viscerotropic Yellow Fever virus, as set forth in SEQ ID NO:1, or any other sequence incorporated herein by reference; b) methods of using a viscerotropic Yellow Fever virus nucleotide sequence for diagnosis of viscerotropic Yellow Fever strains by RT-PCR, gene probes, or expression of antigens c) methods of using the nucleotide sequence of a virulent Yellow Fever virus to identify molecular determinants of viscerotropic disease, in particular using the Hamster as a model system; d) genetic engineering of molecular determinants of viscerotropic phenotypes to improve the safety of live attenuated Yellow Fever vaccines; and e) genetic engineering of the molecular determinants of a virulent phenotype in Yellow Fever virus similar or homologous nucleic acids or proteins in other virus that cause viral hemorrhagic fever. Molecular determinants may include, but are not limited to nucleic acids, polypeptides, complexes of polypeptides, and combinations of thereof. These may not be the same nucleotides/amino acids but could be the same or similar proteins. For example, information derived from Yellow Fever virus may be used to genetically alter dengue viruses, which may help in designing a dengue virus vaccine.

The use of the word "a" or "an" when used in conjunction with the term "comprising" in the claims and/or the specification may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one."

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1 illustrates an exemplary study of the survival of 3-4 week old Golden Syrian hamsters following inoculation with either parental Asibi p0 virus or viscerotropic Asibi p7 virus.

FIG. 2 illustrates an example of viremia in sub-adult hamsters inoculated with either Asibi p7 or Asibi p0 virus. Values shown are the average of 5-6 animals. Downward arrows indicate values that are at or below the limit of sensitivity for this assay.

FIG. 3A-3B shows exemplary H&E stained sections of hamster liver 6 days post infection (dpi). (FIG. 3A) Liver from mock-infected hamster. (FIG. 3B) Liver from hamster infected with Asibi/hamster p0. (FIG. 3C) Liver from hamster infected with Asibi/hamster p7.

FIG. 4A-4B illustrates exemplary liver pathology in hamsters inoculated with viscerotropic Asibi p7. Animal A was sacrificed on day 5 post infection due to severe illness. (Fig. 4A) Steatosis is expressed as a percentage of the total liver. (FIG. 4B) Hepatic necrosis and lobular inflammation are presented as a grade from 0-4 with 0 being none and 4 being severe. The remaining animals (FIG. 4B, animal B-E) were beginning to show signs of illness when they were sacrificed on day 6 post infection (pi).

FIG. 5A-5C illustrates exemplary H&E stained sections of hamster spleen 6 dpi. (FIG. 5A) Spleen from mock-infected hamster. (FIG. 5B) Spleen from hamster infected with Asibi/hamster p0. (FIG. 5C) Spleen from hamster infected with Asibi/hamster p7.

FIG. 6 shows an example of the splenic abnormalities identified in 3-4 week old hamsters inoculated with Asibi p0 and Asibi p7 viruses.

FIG. 7 illustrates the three-dimensional structure of the YF virus E protein based on the crystallographic structure of TBE virus E protein (Rey *et al.*, 1995). The 5 amino

acid positions that differ between the Asibi/hamster p0 and Asibi/hamster p7 E27, E28, E155, E323, and E331 are highlighted and labeled.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Compositions and methods of the present invention include provisions for the improvement of *flavivirus* vaccines so that the risk of disease is reduced or eliminated. In certain embodiments the *flavivirus* is a Yellow Fever virus. In other embodiments, a virus may be an altered 17D, 17D-204, 17DD, or other Yellow Fever vaccines. In various other embodiments the vaccine may be a chimeric vaccine, as described herein. Chimeric refers to a viral genome, viral polypeptide or viral particle that contains a discernable portion(s) of at least two viruses or virus strains, and may also include portions of non-viral nucleic acids and/or polypeptides.

In certain embodiments, viral variants are typically selected that demonstrate an increased virulence in a model host (*e.g.*, a hamster), a so-called supervirulent virus. Supervirulent refers to an organism or virus that demonstrates an increased or enhanced ability to cause injury or disease in a host organism, tissue, and/or cell. Viral isolates may be sequenced to identify nucleotide and/or amino acid changes associated with increased virulence. The information provided by the alterations associated with increased virulence may be used to genetically engineer mutations in other viruses either individually or in various combinations to improve the safety profile of an attenuated virus used as a vaccine. Thus, an engineered virus may then be used as a vaccine with a lower probability of reversion to a virulent phenotype. These alterations will reduce the probability of a reversion in the vaccine by increasing the number of mutational events necessary to alter an encoded amino acid to an amino acid associated with supervirulence or a virulent phenotype.

The information provided by the analysis of nucleotide sequences involved in viscerotropic disease will typically identify nucleotides and amino acids that should not be incorporated in any live attenuated Yellow Fever vaccine and in particular any equivalent position in other *flavivirus* vaccine. An equivalent position may be identified by homology or similarity to Yellow Fever virus sequences or similarity to motifs or

conserved sequence characteristics between a Yellow Fever virus and other members of the *flavivirus* genus.

In an exemplary embodiment, the inventors have passaged a wild-type strain Asibi (Hahn et al., 1987; the parent strain of vaccine strain 17D) seven times in hamsters by liver-to-liver passage and have generated a variant of Asibi virus that is viscerotropic in hamsters, as well as demonstrating a virulent phenotype. Non-hamster passaged Asibi virus does not kill hamsters while Asibi hamster passage 7 virus kills hamsters (supervirulent phenotype).

The genome of Asibi hamster passage 7 (p7) virus has been sequenced and the nucleotide sequence changes associated with the hamster viscerotropic phenotype have been identified by comparing the genomes of non-hamster passaged Asibi virus and Asibi hamster p7 virus. There are a number of nucleotide and amino acid differences between the two viruses. In various embodiments of the invention these mutations may be used to improve 17D, other Yellow Fever virus vaccines, or other *flavivirus* vaccines.

In other embodiments of the invention, genetic engineering may be used to genetically manipulate the single-stranded, positive-sense RNA genome of Yellow Fever virus or other members of the *Flaviviridae* family. Genetic manipulation may introduce mutations into the 17D vaccine virus genome or the genome of another *flavivirus* vaccine virus to further attenuate the virus and reduce the viscerotropic disease potential of 17D. Infectious clones of strain 17D have been developed as a basis for chimeric vaccine (ChimeriVax™) platform to make chimeric 17D viruses containing the foreign envelope protein genes of other *flavivirus*. (*e.g.*, dengue, West Nile and Japanese encephalitis) (Acambis Inc., Cambridge MA). For example see U.S. Patent No. 6,184,024, which is incorporated herein by reference.

In general, the information on the molecular determinants of viscerotropism of Yellow Fever virus is sparse and there is little information regarding the molecular determinants involved in this or other hemorrhagic fevers resulting from *flavivirus* infections. Embodiments of the invention will aid in the identification of these molecular mechanisms and provide for the engineering of improved vaccines.

In certain embodiments, the genomic nucleic acid sequence of Asibi hamster passage 7 virus, as compared with non-hamster passaged Asibi virus (Hahn *et al.*, 1987,

which is incorporated herein by reference), may be used to identify molecular determinants of hamster viscerotropism.

I. *FLAVIVIRUS*

The genus *Flavivirus* is a member of the *Flaviviridae* family and includes the viral subgroups of Yellow Fever virus group, Tick-borne encephalitis virus group, Rio Bravo Group, Japanese encephalitis Group, Tyuleniy Group, Ntaya Group, Uganda S Group, Dengue Group, and Modoc Group. Members of the *Flavivirus* genus may produce a wide variety of disease states, such as fever, arthralgia, rash, hemorrhagic fever, and/or encephalitis. The outcome of infection is influenced by both the virus and host-specific factors, such as age, sex, genetic susceptibility, and/or pre-exposure to the same or a related agent. Some of the various diseases associated with members of the genus *Flavivirus* are Yellow Fever; Dengue Fever; and West Nile, Japanese, and St. Louis Encephalitides.

Virions of the *Flaviviridae* generally contain one molecule of a linear positive-sense single stranded RNA genome of approximately 10,000-11,000 nucleotides that replicates in the cytoplasm of an infected cell. Typically the 5' end of the genome has a cap and the 3' end may or may not have a poly (A) tract. *Flavivirus* are usually transmitted by a vector such as an insect, in many cases the insect is a mosquito.

The viral genome of the *Flavivirus* genus is translated as a single polypeptide and is subsequently cleaved into mature proteins. The proteins encoded by the virus typically consist of structural and non-structural proteins. Generally, there are three structural proteins that typically include the envelope protein (E)(amino acids 286-778 of genbank accession number X03700 and SEQ ID NO:2), the core or capsid protein (C)(amino acids 1-121 of genbank accession number X03700), and the pre-membrane protein (preM)(amino acids 122-285 of genbank accession number X03700)(Hahn *et al.*, 1987). The envelope protein is approximately 493 amino acids with an approximate molecular weight of 50 kDa and is often glycosylated. The envelop protein typically contains twelve conserved cysteine residues which form six disulfide bridges. The core protein is approximately 13 kDa and is rich in arginine and lysine residues. The pre-membrane protein is approximately 10 kDa and is cleaved during or after release of the virus from

infected cells. A cleavage product of the prM protein remains associated with the virion and is approximately 8 kDa and is termed the membrane protein (M). Typically, it is the carboxy terminus of prM that remains associated with the virus particle as the M protein.

The computer databases contain a few entries representative of the Yellow Fever virus genome, which is based on three West African strains and a Trinidad strain. Examples of Genbank entries for representative Yellow Fever virus strains may be found under the following accession numbers: 17D-204 (accession No. X15061), 17D-213 (accession No. U17067), 17DD (accession No. U17066), 17D (accession No. X03700). French viscerotropic virus (accession No. U21056), and French neurotropic virus (accession No. U21055), each of which is incorporated herein by reference. Various other strains or isolates are available in the Genbank, ATCC, or other databases/depositories.

Various members of the *Flaviviridae* family are available through the American Type Culture Collection (Manassas Va.) under the following ATCC numbers: Dengue type 1 (VR-71), Ilheus (VR-73), Japanese encephalitis (VR-74), Murray valley encephalitis (VR-77), Ntaya (VR-78), St Louis encephalitis (VR-80), Uganda S (VR-81), West Nile (VR-82), Zika (VR-84), Dengue type 4 (VR-217), Dengue type 2 (VR-222), Japanese encephalitis (VR-343), Dengue type 1 (VR-344), Dengue type 2 (VR-345), Edge hill (VR-377), Entebbe bat (VR-378), Kokobera (VR-379), Stratford (VR-380), Tembusu (VR-381), Dakar bat (VR-382), Ntaya (VR-78), Banzi (VR-414), Modoc (VR-415), Rio Bravo virus (VR-416), Cowbone ridge (VR-417), Bukalasa (VR-418), Montana myotis leukoencephalitis (VR-537), Bussuquara (VR-557), Sepik (VR-906), Cowbone ridge (VR-1253), Dengue type 2 (VR-1255), Dengue type 3 (VR-1256), Dengue type 4 (VR-1257), Ilheus (VR-1258), Rio Bravo virus (VR-1263), St. Louis encephalitis (VR-1265), West Nile (VR-1267), Dengue type 4 (VR-1490), West Nile (VR-1507), and West Nile (VR-1510), each of which is incorporated herein by reference.

A. Yellow Fever virus

Yellow Fever, as described by the World Health Organization (WHO), is a viral disease that has caused large epidemics in Africa and the Americas. Yellow Fever virus infection causes a wide spectrum of disease, from mild symptoms to severe illness and

death. Although an effective vaccine is available, the number of people infected over the last two decades has increased and Yellow Fever is now a serious public health issue again.

The Yellow Fever virus belongs to the *Flavivirus* genus. In Africa there are five distinct genetic types (called genotypes) associated with East, Central and West Africa (Mutebi *et al.*, 2001). Also, South America has at least two different genotypes.

The virus remains silent in the body during an incubation period of three to six days. There are then two disease phases. While some infections have no symptoms whatsoever, the first, "acute", phase is normally characterized by fever, muscle pain (with prominent backache), headache, shivers, loss of appetite, nausea and/or vomiting. Often, the high fever is paradoxically associated with a slow pulse. After three to four days most patients improve and their symptoms disappear.

However, 15% enter a "toxic phase" within 24 hours. Fever reappears and several body systems are affected. The patient rapidly develops jaundice and complains of abdominal pain with vomiting. Bleeding can occur from the mouth, nose, eyes and/or stomach. Once this happens, blood appears in the vomit and feces. Kidney function deteriorates; this can range from abnormal protein levels in the urine (albuminuria) to complete kidney failure with no urine production (anuria). Up to half of the patients in the "toxic phase" die within 10-14 days. The remainder recover without significant organ damage.

Yellow Fever is difficult to recognize, especially during the early stages. It can easily be confused with malaria, typhoid, rickettsial diseases, hemorrhagic viral fevers (e.g. Lassa), arboviral infections (e.g. dengue), leptospirosis, viral hepatitis and poisoning (e.g. carbon tetrachloride). A laboratory analysis is required to confirm a suspected case. Blood tests (serology assays) can detect Yellow Fever antibodies that are produced in response to the infection. Several other techniques are used to identify the virus itself in blood specimens or liver tissue collected after death.

B. Flaviviral Nucleic Acid Compositions

The present invention concerns *flaviviruses* that are advantageous in the study and treatment of a variety of diseases. It concerns *flaviviruses*, particularly Yellow Fever

viruses, that have been either derived from serial passage in a model host organism, such as a hamster, or constructed with one or more nucleotide alterations compared to wild-type or vaccine strains, such that the virus has desirable properties for use against viral infection, while being less likely to revert to a virulent phenotype. The teachings described herein provide various methods, by way of example, of implementing methods and compositions of the invention. They provide background for generating altered or mutant viruses through the use of propagation in a model host, as well as the genetic engineering of viruses to reduce the probability of reversion to a virulent phenotype. Genetic engineering may include various known methods of manipulating nucleic acid to produce a desired nucleic acid sequence (see Sambrook *et al.*, 1989)

In certain embodiments, the present invention concerns generating a Yellow Fever virus with an altered phenotype, for example a virus that is more virulent than a parental form of the virus; an example of a parental strain is the Asibi strain of Yellow Fever virus. In other embodiments, the present invention concerns analyzing the resultant more virulent virus(es) and using this information to engineer an improved strain of virus for vaccination. This improved strain of virus may be used in combination with proteinaceous compositions as part of a pharmaceutically acceptable formulation. Compositions of the invention may be used as a vaccine to vaccinate an organism against Yellow Fever virus infection

C. Nucleic Acid Molecules

1. Polynucleotides Encoding Native Proteins or Modified Proteins

The present invention concerns polynucleotides, isolatable from cells or virions, that are capable of expressing all or part of a protein, polypeptide, and/or virus. In some embodiments of the invention, it concerns a viral genome that has been specifically mutated to generate a virus with a virulent phenotype or an improved characteristic or property, *e.g.*, a reduced probability of reversion. The polynucleotides may encode a peptide, polypeptide, and/or virus containing all or part of a viral amino acid sequence or they may be engineered so they do not encode such a viral polypeptide or encode a viral polypeptide having at least one function or activity reduced, diminished, or absent. The

polynucleotides may comprise a chimeric virus, a virus derived from genetic material of two separate viruses.

As used herein, the term "nucleic acid segment" refers to a nucleic acid molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a nucleic acid segment encoding a polypeptide refers to a nucleic acid segment that contains wild-type, polymorphic, or mutant polypeptide-coding sequences yet is isolated away from, or purified frée from, total mammalian or human genomic DNA. Included within the term "nucleic acid segment" are a polynucleotide or polynucleotides, nucleic acid segments smaller than a polynucleotide, and recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like.

As used in this application, the term "*flavivirus* polynucleotide or nucleic acid" refers to a nucleic acid molecule encoding a *flavivirus* or a flaviviral polypeptide that has been isolated free of total genomic nucleic acid. Similarly, a "Yellow Fever virus polynucleotide or nucleic acid" refers to a nucleic acid molecule encoding a Yellow Fever virus or a Yellow Fever viral polypeptide that has been isolated free of total genomic nucleic acid. A "*flavivirus* genome" or a "Yellow Fever virus genome" refers to a nucleic acid molecule that can be provided to a host cell to yield a viral particle, in the presence or absence of a helper virus. The genome may or may have not been genetically altered as compared to wild-type virus.

The term "cDNA" is intended to refer to DNA prepared using messenger RNA (mRNA) or RNA encoding polypeptides as a template. The advantage of using a cDNA, as opposed to genomic DNA or DNA polymerized from a genomic, non- or partially-processed RNA template, is that the cDNA primarily contains coding sequences of the corresponding protein.

It also is contemplated that a particular polypeptide from a given species may be represented by natural variants that have slightly different nucleic acid sequences but, nonetheless, encode the same protein (see Table 1).

Similarly, a polynucleotide comprising an isolated or purified wild-type or mutant gene refers to a nucleic acid segment including wild-type or mutant polypeptide coding sequences and, in certain aspects, regulatory sequences, isolated substantially away from other naturally occurring genes or protein encoding sequences. In this respect, the term

"gene" is used for simplicity to refer to a functional protein, polypeptide, or peptide-encoding unit (including any sequences required for proper transcription, post-translational modification, or localization). As will be understood by those in the art, this functional term includes genomic sequences, positive strand RNA, cDNA sequences, and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins, and mutants. A nucleic acid encoding all or part of a native or modified polypeptide may contain a contiguous nucleic acid sequence encoding all or a portion of such a polypeptide of the following lengths: 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 441, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1095, 1100, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 9000, 10000, 10,862, 11,000 or more nucleotides, nucleosides, or base pairs.

In particular embodiments, the invention concerns isolated nucleic acid segments and recombinant vectors incorporating nucleic acid sequences that encode a wild-type or mutant *flavivirus*, in particular Yellow Fever virus, polypeptide or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially corresponding to a native polypeptide. Thus, an isolated nucleic acid segment or vector containing a nucleic acid segment may encode, for example, an envelope protein. The term "recombinant" may be used in conjunction with a polypeptide or the name of a specific polypeptide, and this generally refers to a polypeptide produced from a nucleic acid molecule that has been manipulated *in vitro, in situ* or that is the replicated product of such a molecule.

In other embodiments, the invention concerns isolated nucleic acid segments and recombinant vectors incorporating nucleic acid sequences that encode a polypeptide or peptide that includes within its amino acid sequence a contiguous amino acid sequence in accordance with, or essentially corresponding to the polypeptide.

The nucleic acid segments used in the present invention, regardless of the length of the coding sequence itself, may be combined with other nucleic acid sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

It is contemplated that the nucleic acid constructs of the present invention may encode full-length polypeptide from any source or encode a truncated version of the polypeptide, for example a truncated Yellow Fever virus polypeptide, such that the transcript of the coding region represents the truncated version. The truncated transcript may then be translated into a truncated protein. Alternatively, a nucleic acid sequence may encode a full-length polypeptide sequence with additional heterologous coding sequences, for example to allow for purification of the polypeptide, transport, secretion, post-translational modification, protease cleavage or for therapeutic benefits such as targeting, antigenicity or efficacy. As discussed above, a tag or other heterologous polypeptide may be added to the modified polypeptide-encoding sequence, wherein "heterologous" refers to a polypeptide that is not the same as the modified polypeptide.

In a non-limiting example, one or more nucleic acid constructs may be prepared that include a contiguous stretch of nucleotides identical to or complementary to the a particular gene or segment of a viral genome, such as the envelope protein gene. A nucleic acid construct may be at least 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 15,000, 20,000, 30,000, 50,000, 100,000, 250,000, 500,000, 750,000, to at least 1,000,000 nucleotides in length, as well as constructs of greater size, up to and including chromosomal sizes (including all intermediate lengths and intermediate ranges), given the advent of nucleic acids constructs such as a yeast artificial chromosome are known to those of ordinary skill in the art. It will be readily understood that "intermediate lengths" and "intermediate ranges," as used herein, means any length or range including or between the quoted values (*i.e.*, all integers including and between such values).

The nucleic acid segments used in the present invention encompass biologically functional equivalent modified polypeptides and peptides, for example, a modified envelope protein. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by a human may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements in reversion frequency of a virus, in antigenicity of a protein, or in the efficacy of any treatment or vaccine involving the protein or virus.

In certain embodiments, the invention concerns isolated nucleic acids, nucleic acid segments and recombinant vectors that include within their sequence a contiguous nucleic acid sequence from that shown in SEQ ID NO:1, or any other sequence incorporated by reference. Such sequences, however, may be mutated to yield a virus that is altered with respect to a wild-type or a vaccine strain of a virus, *e.g.*, Yellow Fever virus or its vaccine derivatives.

It also will be understood that this invention is not limited to the particular nucleic acid and amino acid sequences of SEQ ID NO:1, 2, 3, 4 or any other sequence incorporated by reference. Recombinant vectors and isolated nucleic acid segments may therefore variously include the Yellow Fever virus-coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region or codons, or they may encode larger polypeptides that nevertheless include viral-coding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acid sequences.

The nucleic acid segments of the present invention encompass biologically functional equivalent Yellow Fever virus proteins and peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure

may be engineered, based on considerations of the properties of the amino acids being exchanged or their representative codons. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the virus resulting in a reduced probability of reversion to a virulent phenotype.

2. Mutagenesis of Flaviviral Polynucleotides

Where employed, mutagenesis will be accomplished by a variety of standard, mutagenic procedures, including passaging virus through cell lines or animals, and standard molecular biological techniques, for exemplary methods see Tech *et al.* 2001 and Sambrook *et al.*, 1989. Mutation is the process whereby changes occur in the quantity or structure of a nucleic acid, a polypeptide, or an organism. Mutation can involve modification of a single nucleotide, the nucleotide sequence of a single gene, blocks of genes or whole chromosomes or genomes. Changes in single genes may be the consequence of point mutations which involve the removal, addition or substitution of a single nucleotide base within a nucleic acid sequence, or they may be the consequence of changes involving the insertion or deletion of large numbers of nucleotides.

Mutations may be induced following exposure to chemical or physical mutagens. Such mutation-inducing agents include ionizing radiation, ultraviolet light (U.V.) and a diverse array of chemicals such as alkylating agents and polycyclic aromatic hydrocarbons all of which are capable of interacting either directly or indirectly (generally following some metabolic biotransformations) with nucleic acids. The DNA damage induced by such agents may lead to modifications of base sequence when the affected DNA is replicated or repaired and thus to a mutation. Mutation also can be site-directed through the use of particular targeting methods, such as oligo directed site directed mutatgenesis.

a. Random Mutagenesis

i) Insertional Mutagenesis

Insertional mutagenesis is based on the inactivation of a gene via insertion of a known DNA fragment. Because it involves the insertion of some type of DNA fragment,

the mutations generated are generally loss-of-function, rather than gain-of-function mutations. However, there are several examples of insertions generating gain-of-function mutations (Oppenheimer et al. 1991). Insertion mutagenesis has been very successful in bacteria and Drosophila (Cooley et al. 1988) and recently has become a powerful tool in corn (Schmidt et al. 1987); Arabidopsis; (Marks et al., 1991; Koncz et al. 1990); and Antirrhinum (Sommer et al. 1990). Insertional mutagenesis may be accomplished using standard molecular biology techniques.

ii) Chemical mutagenesis

Chemical mutagenesis offers certain advantages, such as the ability to find a full range of mutations with degrees of phenotypic severity, and is facile and inexpensive to perform. The majority of chemical carcinogens produce mutations in DNA. Benzo[a]pyrene, N-acetoxy-2-acetyl aminofluorene and aflotoxin B1 cause GC to TA transversions in bacteria and mammalian cells. Benzo[a]pyrene also can produce base substitutions such as AT to TA. N-nitroso compounds produce GC to AT transitions. Alkylation of the O4 position of thymine induced by exposure to n-nitrosoureas results in TA to CG transitions.

iii) Radiation Mutagenesis

Biological molecules are degraded by ionizing radiation. Adsorption of the incident energy leads to the formation of ions and free radicals, and breakage of some covalent bonds. Susceptibility to radiation damage appears quite variable between molecules, and between different crystalline forms of the same molecule. It depends on the total accumulated dose, and also on the dose rate (as once free radicals are present, the molecular damage they cause depends on their natural diffusion rate and thus upon real time). Damage is reduced and controlled by making the sample as cold as possible. Ionizing radiation causes DNA damage, generally proportional to the dose rate.

In the present invention, the term "ionizing radiation" means radiation comprising particles or photons that have sufficient energy or can produce sufficient energy via nuclear interactions to produce ionization (gain or loss of electrons). An exemplary and preferred ionizing radiation is an x-radiation. The amount of ionizing radiation needed in a given cell generally depends upon the nature of that cell and the nature of the mutation target. Means for determining an effective amount of radiation are well known in the art.

iv) In Vitro Scanning Mutagenesis

Random mutagenesis also may be introduced using error prone PCR (Cadwell and Joyce, 1992). The rate of mutagenesis may be increased by performing PCR in multiple tubes with dilutions of templates.

One particularly useful mutagenesis technique is alanine scanning mutagenesis in which a number of residues are substituted individually with the amino acid alanine so that the effects of losing side-chain interactions can be determined, while minimizing the risk of large-scale perturbations in protein conformation (Cunningham et al., 1989).

In vitro scanning saturation mutagenesis provides a rapid method for obtaining a large amount of structure-function information including: (i) identification of residues that modulate ligand binding specificity, (ii) a better understanding of ligand binding based on the identification of those amino acids that retain activity and those that abolish activity at a given location, (iii) an evaluation of the overall plasticity of an active site or protein subdomain, (iv) identification of amino acid substitutions that result in increased binding.

b. Site-Directed Mutagenesis

Structure-guided site-specific mutagenesis represents a powerful tool for the dissection and engineering of proteins. The technique provides for the preparation and testing of sequence variants by introducing one or more nucleotide sequence changes into a selected DNA.

Site-specific mutagenesis uses specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent, unmodified nucleotides. In this way, a primer sequence is provided with sufficient size and complexity to form a stable duplex on both sides of the deletion junction being traversed. A primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

The technique typically employs a bacteriophage vector that exists in both a single-stranded and double-stranded form. Vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double-stranded

plasmids are also routinely employed in site-directed mutagenesis, which eliminates the step of transferring the gene of interest from a phage to a plasmid.

In general, one first obtains a single-stranded vector, or melts two strands of a double-stranded vector, which includes within its sequence a DNA sequence encoding the desired protein or genetic element. An oligonucleotide primer bearing the desired mutated sequence, synthetically prepared, is then annealed with the single-stranded DNA preparation, taking into account the degree of mismatch when selecting hybridization conditions. The hybridized product is subjected to DNA polymerizing enzymes such as E. coli polymerase I (Klenow fragment) in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed, wherein one strand encodes the original non-mutated sequence, and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate host cells, such as E. coli cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement.

Comprehensive information on the functional significance and information content of a given residue of protein can best be obtained by saturation mutagenesis in which all 19 amino acid substitutions are examined. The shortcoming of this approach is that the logistics of multiresidue saturation mutagenesis are daunting (Warren et al., 1996, Brown et al., 1996; Zeng et al., 1996; Burton and Barbas, 1994; Yelton et al., 1995; Jackson et al., 1995; Short et al., 1995; Wong et al., 1996; Hilton et al., 1996). Hundreds, and possibly even thousands, of site specific mutants must be studied. However, improved techniques make production and rapid screening of mutants much more straightforward. See also, U.S. Patents 5,798,208 and 5,830,650, for a description of "walk-through" mutagenesis. Other methods of site-directed mutagenesis are disclosed in U.S. Patents 5,220,007; 5,284,760; 5,354,670; 5,366,878; 5,389,514; 5,635,377; and 5,789,166.

D. Oligonucleotide Probes and Primers

Naturally, the present invention also encompasses nucleic acid segments that are complementary, or essentially complementary, to all or part of the sequence set forth in SEQ ID NO:1, or any other sequence incorporated by reference. Nucleic acid sequences that are

"complementary" are those that are capable of base-pairing according to the standard Watson-Crick complementary rules. As used herein, the term "complementary sequences" means nucleic acid sequences that are substantially complementary, as may be assessed by the same nucleotide comparison set forth above, or as defined as being capable of hybridizing to the nucleic acid segment of SEQ ID NO:1, or any other sequence incorporated by reference, under relatively stringent conditions such as those described herein. Such sequences may encode the entire sequence of *flavivirus* genome or functional or non-functional fragments thereof.

Alternatively, the hybridizing segments may be shorter oligonucleotides. Sequences of 17 bases long should occur only once in the human genome and, therefore, suffice to specify a unique target sequence in the presence of various nucleic acids. Although shorter oligomers are easier to make and increase *in vivo* accessibility, numerous other factors are involved in determining the specificity of hybridization. Both binding affinity and sequence specificity of an oligonucleotide to its complementary target increases with increasing length. It is contemplated that exemplary oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250 or more base pairs will be used, although others are contemplated. Longer polynucleotides encoding 250, 500, 1000, 1212, 1500, 2000, 2500, 3000 or 3431 bases and longer are contemplated as well. Such oligonucleotides will find use, for example, as probes in Southern and RNA blots and as primers in nucleic acid amplification reactions.

Suitable hybridization conditions will be well known to those of skill in the art. In certain applications, for example, substitution of amino acids by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, a medium stringency condition could be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition could be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Thus, hybridization conditions can be readily manipulated and thus will generally be a method of choice depending on the desired results.

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, at temperatures ranging from approximately 40°C to about 72°C. Formamide and SDS also may be used to alter the hybridization conditions.

One method of using probes and primers of the present invention is in the search for other viral sequences related to Yellow Fever virus or, more particularly, homologs of the envelope protein or other yellow virus protein sequences. By varying the stringency of hybridization, and the region of the probe, different degrees of homology may be discovered.

Another way of exploiting probes and primers of the present invention is in site-directed, or site-specific, mutagenesis. The technique provides a ready ability to prepare and test sequence variants, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into complementary nucleic acid. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the nucleic acid sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion or mutation junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

The technique typically employs a bacteriophage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage vectors are commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis, which eliminates the step of transferring the nucleic acid of interest from a phage to a plasmid.

In general, site-directed mutagenesis is performed by first obtaining a single-stranded vector, or melting of two strands of a double stranded vector which includes

within its sequence a nucleic acid sequence encoding the desired protein or protein segment, protein segment being any part or fragment of an encoded protein. An oligonucleotide primer bearing the desired mutated sequence is synthetically prepared. This primer is then annealed with the single-stranded nucleic acid preparation, taking into account the degree of mismatch when selecting hybridization conditions, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected that include recombinant vectors bearing the mutated sequence arrangement. There are newer and simpler site-directed mutagenesis techniques that can also be employed for this purpose. These include procedures marketed in kit form that are readily available to one of ordinary skill in the art.

The preparation of sequence variants of the selected nucleic acid using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting, as there are other ways in which sequence variants of nucleic acids may be obtained. For example, recombinant vectors encoding the desired nucleic acid segment may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

E. Proteinaceous Compositions

Embodiments of the invention may include viral particles, including proteins and polypeptides associated with *flavivirus* particles. In various embodiments the viral particles may be produced and/or propagated from an altered nucleic acid encoding a *flavivirus*, in particular a Yellow Fever virus. In certain embodiments the altered nucleic acid encodes a virus with enhanced virulence. In other embodiments the nucleic acid may be engineered to encode a virus with a reduced probability of reverting to a virulent phenotype. As used herein, a "protein" or "polypeptide" refers to a molecule comprising at least one amino acid residue. In some embodiments, a wild-type version of a protein or polypeptide may be employed, however, in many embodiments of the invention, a

viral protein or polypeptide is absent or altered so as to render the virus more useful for the treatment of a subject or patient. The terms described above may be used interchangeably herein. A "modified protein" or "modified polypeptide" refers to a protein or polypeptide whose chemical structure is altered with respect to the wild-type or parental (*i.e.*, a *flavivirus* polynucleotide to be altered, which may be a vaccine strain and not considered wild-type) protein or polypeptide. In some embodiments, a modified protein or polypeptide has at least one modified activity or function (recognizing that proteins or polypeptides may have multiple activities or functions). The modified activity or function may be reduced, diminished, eliminated, enhanced, improved, or altered in some other way (such as specificity or propensity to revert to a virulent phenotype) with respect to that activity or function in a wild-type or vaccine protein or polypeptide. It is specifically contemplated that a modified protein or polypeptide may be altered with respect to one activity or function yet retain wild-type or vaccine activity or function in other respects. All or part of a *flavivirus* encoded protein may be isolated using known recombinant techniques and used as part of proteinaceous composition, *e.g.*, as a peptide vaccine or to generate *flavivirus* specific antibodies.

In certain embodiments the size of a protein or polypeptide may comprise, but is not limited to, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500 or greater amino molecule residues, and any range derivable therein.

As used herein, an "amino molecule" refers to any amino acid, amino acid derivative or amino acid mimic as would be known to one of ordinary skill in the art. In certain embodiments, the residues of the proteinaceous molecule are sequential, without any non-amino molecule interrupting the sequence of amino molecule residues. In other embodiments, the sequence may comprise one or more non-amino molecule moieties. In

particular embodiments, the sequence of residues of the proteinaceous molecule may be interrupted by one or more non-amino molecule moieties.

Accordingly, the term "proteinaceous composition" encompasses amino molecule sequences comprising at least one of the 20 common amino acids in naturally synthesized proteins, or at least one modified or unusual amino acid.

In certain embodiments the proteinaceous composition comprises at least one protein, polypeptide or peptide. In further embodiments the proteinaceous composition comprises a biocompatible protein, polypeptide or peptide. As used herein, the term "biocompatible" refers to a substance that produces no significant untoward effects when applied to, or administered to, a given organism according to the methods and amounts described herein. Such untoward or undesirable effects are those such as significant toxicity or adverse immunological reactions. In preferred embodiments, biocompatible protein, polypeptide or peptide containing compositions will generally be essentially free from toxins, pathogens and harmful immunogens.

Proteinaceous compositions may be made by any technique known to those of skill in the art, including the expression of proteins, polypeptides or peptides through standard molecular biological techniques, the isolation of proteinaceous compounds from natural sources, or the chemical synthesis of proteinaceous materials.

In certain embodiments a proteinaceous compound may be purified. Generally, "purified" will refer to a specific protein, polypeptide, or peptide composition that has been subjected to fractionation to remove various other proteins, polypeptides, or peptides, and which composition substantially retains its activity, as may be assessed, for example, by the protein assays, as would be known to one of ordinary skill in the art for the specific or desired protein, polypeptide or peptide.

1. Variants of Viral Polypeptides

Alteration in the nucleic acids encoding a *flavivirus* may be altered so that the probability of a virus reverting to a virulent phenotype is reduced. Nucleic acid alteration(s) may include the substitution of an amino acid in a vaccine strain with a conservative or non-conservative amino acid, so that multiple mutations are needed to change an amino acid in a vaccine or other virus strain to an amino acid present in a virulent virus.

Amino acid sequence variants of the polypeptides of the present invention can be substitutional, insertional or deletion variants. A mutation in a gene encoding a viral polypeptide may affect 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 or more non-contiguous or contiguous amino acids of the polypeptide, as compared to wild-type.

Deletion variants lack one or more residues of the parental, native or wild-type protein. Individual residues can be deleted or all or part of a domain (such as a catalytic or binding domain) can be deleted. Insertional mutants typically involve the addition of material at a non-terminal point in the polypeptide. This may include the insertion of an immunoreactive epitope or simply one or more residues. Terminal additions, called fusion proteins, may also be generated.

In certain embodiments, substitutions will be made so that multiple mutations in a codon will be necessary to encode for a amino acid that is associated with increased virulence. Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide, with or without the loss of other functions or properties. Substitutions may be conservative, that is, one amino acid is replaced with one of similar shape and charge. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or

valine; leucine to valine or isoleucine; lysine to arginine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine. Alternatively, substitutions may be non-conservative such that a function or activity of the polypeptide is affected or is not affected. Non-conservative changes typically involve substituting a residue with one that is chemically dissimilar, such as a polar or charged amino acid for a nonpolar or uncharged amino acid, and vice versa.

The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six codons for arginine or serine, and also refers to codons that encode biologically equivalent amino acids (see Table 1, below).

TABLE 1

Codon Table

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

It also will be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences,

and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.

The following is a discussion based upon changing of the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, cellular receptors or binding sites on target or immune effector cells. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid substitutions can be made in a protein sequence, and in its underlying coding sequence, and nevertheless produce a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the DNA sequences of genes without appreciable loss of their biological utility or activity and still result in a vaccine with a reduced probability of reversion to a virulent form of *flavivirus*. Table 1 shows the codons that encode particular amino acids.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte & Doolittle, 1982). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

It also is understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein. As detailed in U.S. Patent 4,554,101, the following

hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine *-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still produce a biologically equivalent and immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions generally are based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take into consideration the various foregoing characteristics are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

II. METHODS OF DETECTION

In various embodiments, the detection of *flavivirus*, in particular Yellow Fever virus, may be used to identify infection by a virulent form of the virus or to confirm the identity of a particular vaccine or strain. Detection methods may use the antigenic properties of a virus particle or the properties of the nucleic acid component of the virus to identify and/or detect the presence of a virus.

A. Nucleic Acid Detection

In addition to their use in directing the expression of *flavivirus* proteins, polypeptides and/or peptides, the nucleic acid sequences disclosed herein have a variety of other uses. For example, they have utility as probes or primers for embodiments involving nucleic acid hybridization or amplification. They may be used in diagnostic or screening methods of the present invention. Detection of nucleic acids encoding *flavivirus* or *flavivirus* polypeptide modulators are encompassed by the invention.

1. Hybridization

The use of a probe or primer of between 13 and 100 nucleotides, preferably between 17 and 100 nucleotides in length, or in some aspects of the invention up to 1-2 kilobases or more in length, allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over contiguous stretches greater than 20 bases in length are generally preferred, to increase stability and/or selectivity of the hybrid molecules obtained. One will generally prefer to design nucleic acid molecules for hybridization having one or more complementary sequences of 20 to 30 nucleotides, or even longer where desired. Such fragments may be readily prepared, for example, by directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

Accordingly, the nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of DNAs and/or RNAs or to provide primers for amplification of DNA or RNA from samples. Depending on the application envisioned, one would desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of the probe or primers for the target sequence.

For applications requiring high selectivity, one will typically desire to employ relatively high stringency conditions to form the hybrids. For example, relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50°C to about 70°C. Such high stringency conditions tolerate little, if any, mismatch between the probe or primers and the template or target strand and would be particularly suitable for isolating specific nucleic acids or for detecting specific RNA transcripts. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide.

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 1.0 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized could include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, at temperatures ranging from approximately 40°C to about 72°C.

In certain embodiments, it will be advantageous to employ nucleic acids of defined sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may desire to employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a detection means that is visibly or spectrophotometrically detectable, to identify specific hybridization with complementary nucleic acid containing samples.

In general, it is envisioned that the probes or primers described herein will be useful as reagents in solution hybridization, as in PCR™, for detection of expression of corresponding nucleic acids, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under desired conditions. The conditions selected will depend on the particular circumstances (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Optimization of hybridization conditions for the particular application of interest is well known to those of skill in the art. After washing of the hybridized molecules to remove non-specifically bound probe molecules, hybridization is detected, and/or quantified, by determining the amount of bound label. Representative solid phase hybridization methods are disclosed in U.S. Patents 5,843,663, 5,900,481 and 5,919,626, each of which is incorporated herein by reference. Other methods of hybridization that may be used in the practice of the present invention are disclosed in U.S. Patents 5,849,481, 5,849,486 and 5,851,772, also incorporated herein by reference.

2. Amplification of Nucleic Acids

Nucleic acids used as a template for amplification may be isolated from cells, tissues, viral isolates, blood or other samples according to standard methodologies (Sambrook *et al.*, 1989). In certain embodiments, analysis is performed on whole cell or

tissue homogenates or biological fluid samples without substantial purification of the template nucleic acid. The nucleic acid may be genomic DNA, viral RNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to first convert the RNA to a complementary DNA.

The term "primer," as used herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty and/or thirty base pairs in length, but longer sequences can be employed. Primers may be provided in double-stranded and/or single-stranded form, although the single-stranded form is preferred.

Pairs of primers designed to selectively hybridize to nucleic acids corresponding to SEQ ID NO:1, or any other sequence incorporated by reference, or any other segment thereof corresponding to a nucleic acid sequence are contacted with the template nucleic acid under conditions that permit selective hybridization. Depending upon the desired application, high stringency hybridization conditions may be selected that will only allow hybridization to sequences that are completely complementary to the primers. In other embodiments, hybridization may occur under reduced stringency to allow for amplification of nucleic acids contain one or more mismatches with the primer sequences. Once hybridized, the template-primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

The amplification product may be detected or quantified. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent label or even via a system using electrical and/or thermal impulse signals (Bellus, 1994).

A number of template dependent processes are available to amplify the oligonucleotide sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR^{TM}) which is described in detail in U.S. Patents 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*,

1988, each of which is incorporated herein by reference in their entirety. Polymerase chain reaction methodologies are well known in the art.

Another method for amplification is ligase chain reaction ("LCR"), disclosed in European Application No. 320 308, incorporated herein by reference in its entirety. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence. A method based on PCR™ and oligonucleotide ligase assay (OLA), disclosed in U.S. Patent 5,912,148, may also be used.

Alternative methods for amplification of target nucleic acid sequences that may be used in the practice of the present invention are disclosed in U.S. Patents 5,843,650, 5,846,709, 5,846,783, 5,849,546, 5,849,497, 5,849,547, 5,858,652, 5,866,366, 5,916,776, 5,922,574, 5,928,905, 5,928,906, 5,932,451, 5,935,825, 5,939,291 and 5,942,391, GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety. Qbeta Replicase, described in PCT Application No. PCT/US87/00880, may also be used as an amplification method in the present invention.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention (Walker *et al.*, 1992). Strand Displacement Amplification (SDA), disclosed in U.S. Patent 5,916,779, is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (Kwoh *et al.*, 1989; PCT Application WO 88/10315, incorporated herein by reference in their entirety). European Application No. 329 822 disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention.

PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a

promoter region/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" and "one-sided PCR" (Frohman, 1990; Ohara *et al.*, 1989).

3. Detection of Nucleic Acids

Following any amplification, it may be desirable to separate the amplification product from the template and/or the excess primer. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using standard methods (Sambrook *et al.*, 1989).

In certain embodiments, the amplification products are visualized. A typical visualization method involves staining of a gel with ethidium bromide and visualization of bands under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the separated amplification products can be exposed to x-ray film or visualized under the appropriate excitatory spectra.

In particular embodiments, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art (see Sambrook *et al.*, 1989). One example of the foregoing is described in U.S. Patent 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

Other methods of nucleic acid detection that may be used in the practice of the instant invention are disclosed in U.S. Patents 5,840,873, 5,843,640, 5,843,651, 5,846,708, 5,846,717, 5,846,726, 5,846,729, 5,849,487, 5,853,990, 5,853,992, 5,853,993, 5,856,092, 5,861,244, 5,863,732, 5,863,753, 5,866,331, 5,905,024, 5,910,407, 5,912,124, 5,912,145, 5,919,630, 5,925,517, 5,928,862, 5,928,869, 5,929,227, 5,932,413 and 5,935,791, each of which is incorporated herein by reference.

4. Other Assays

Other methods for genetic screening may be used within the scope of the present invention, for example, to detect mutations in genomic RNA, cDNA and/or RNA samples. Methods used to detect point mutations include denaturing gradient gel electrophoresis ("DGGE"), restriction fragment length polymorphism analysis ("RFLP"), chemical or enzymatic cleavage methods, direct sequencing of target regions amplified by PCR™ (see above), single-strand conformation polymorphism analysis ("SSCP") and other methods well known in the art.

One method of screening for point mutations is based on RNase cleavage of base pair mismatches in RNA/DNA or RNA/RNA heteroduplexes. As used herein, the term "mismatch" is defined as a region of one or more unpaired or mispaired nucleotides in a double-stranded RNA/RNA, RNA/DNA or DNA/DNA molecule. This definition thus includes mismatches due to insertion/deletion mutations, as well as single or multiple base point mutations.

U.S. Patent 4,946,773 describes an RNase A mismatch cleavage assay that involves annealing single-stranded DNA or RNA test samples to an RNA probe, and subsequent treatment of the nucleic acid duplexes with RNase A. For the detection of mismatches, the single-stranded products of the RNase A treatment, electrophoretically separated according to size, are compared to similarly treated control duplexes. Samples containing smaller fragments (cleavage products) not seen in the control duplex are scored as positive.

Other investigators have described the use of RNase I in mismatch assays. The use of RNase I for mismatch detection is described in literature from Promega Biotech. Promega markets a kit containing RNase I that is reported to cleave three out of four known mismatches. Others have described using the MutS protein or other DNA-repair enzymes for detection of single-base mismatches.

Alternative methods for detection of deletion, insertion or substitution mutations that may be used in the practice of the present invention are disclosed in U.S. Patents 5,849,483, 5,851,770, 5,866,337, 5,925,525 and 5,928,870, each of which is incorporated herein by reference in its entirety.

B. Protein Detection

In various embodiments, *Flavivirus*, in particular Yellow Fever virus, may be detected by using polyclonal or monoclonal antibodies in standard immunochemical procedures, such as ELISA and Western blot methods and in immunohistochemical procedures such as tissue staining, as well as in other procedures which may utilize antibodies specific to *flavivirus*-related antigen epitopes. For general methodologies regarding antibody production and utilization see Harlow and Lane, 1988; and Sambrook *et al.*, 1989, each of which is incorporated herein by reference.

III. PHARMACEUTICAL FORMULATIONS

In various embodiments of the present invention, a method of treatment or prophylaxis for a viral infection is contemplated. Examples of viral infection contemplated for treatment include Yellow Fever virus, Japanese encephalitis virus, Dengue fever virus, West Nile virus, hepatitis C virus, St. Louis encephalitis virus, and other members of the *flavivirus* genus described herein may be treated. Vaccines of the invention may be suitable to induce an immune response against a *flavivirus*, Yellow Fever virus or a derivative thereof. See U.S. Patent Nos. 6,372,221, 6,337,073, 6,254,873, 6,184,024, 6,171,854, 5,744,141, 5,744,140, 5,736,148, 4,810,492, and 4,500,512, each incorporated herein by reference, for exemplary methods and compositions related to *flavivirus* and there use in vaccines.

An exemplary vaccine composition may include a Yellow Fever virus with a viral genome with at least one of the following alterations: a) an alteration in the nucleic acid sequence encoding amino acid 323 of an/the envelope protein, wherein the first alteration requires more than one nucleotide change to encode an arginine; b) an alteration in the nucleic acid sequence encoding amino acid 27 of an/the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a histidine; etc., c) an alteration in the nucleic acid sequence encoding amino acid 28 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a glycine; d) an alteration in the nucleic acid sequence encoding amino acid 155 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a alanine; e) an alteration in the nucleic acid sequence encoding amino

acid 331 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a arginine; f) an alteration in the nucleic acid sequence encoding amino acid 48 of the NS2A protein, wherein the second alteration requires more than one nucleotide change to encode a alanine; or g) an alteration in the nucleic acid sequence encoding amino acid 98 of the NS4B protein, wherein the second alteration requires more than one nucleotide change to encode a isoleucine. In other embodiments the viral genome may include one, two, three, four, five, six, or seven of the above alterations. In yet other embodiments, the vaccine compositions described herein may be used in methods of vaccination that include administering the vaccine compositions to a subject in need of vaccination. Each of these alteration may be used in conjunction with any other combination of alteration. Such that any one alteration may be used in combination with one, two, three, four, five, or six of the other alterations described herein.

An effective amount of the pharmaceutical composition, generally, is defined as that amount sufficient to detectably and repeatedly ameliorate, reduce, minimize or limit the extent of the infection, disease or its symptoms. More rigorous definitions may apply, including elimination, eradication or cure of disease.

Pharmaceutical compositions of the present invention comprise an effective amount of one or more attenuated virus of the *Flaviviridae* family with a mutant or altered viral genome and/or additional agent(s) dissolved or dispersed in a pharmaceutically acceptable carrier. The phrases "pharmaceutical or pharmacologically acceptable" refers to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal, such as, for example, a human, as appropriate. The preparation of a pharmaceutical composition that contains at least one attenuated virus of the *Flaviviridae* family with a mutant or altered viral genome and/or additional agent(s) dissolved or dispersed in a pharmaceutically acceptable carrier will be known to those of skill in the art in light of the present disclosure, as exemplified by Remington's Pharmaceutical Sciences, 18th Ed. Mack Printing Company, 1990, incorporated herein by reference. Moreover, for animal (e.g., human) administration, it will be understood that preparations should meet sterility,

pyrogenicity, general safety and purity standards as required by FDA Office of Biological Standards.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, surfactants, antioxidants, preservatives (e.g., antibacterial agents, antifungal agents), isotonic agents, absorption delaying agents, salts, preservatives, drugs, drug stabilizers, gels, binders, excipients, disintegration agents, lubricants, sweetening agents, flavoring agents, dyes, such like materials and combinations thereof, as would be known to one of ordinary skill in the art (see, for example, Remington's Pharmaceutical Sciences, 1990, incorporated herein by reference). Except insofar as any conventional carrier is incompatible with the active ingredient, its use in the therapeutic or pharmaceutical compositions is contemplated.

The attenuated virus of the invention may be formulated into a composition in a free base, neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts, e.g., those formed with the free amino groups of a proteinaceous composition, or which are formed with inorganic acids such as for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric or mandelic acid. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as for example, sodium, potassium, ammonium, calcium or ferric hydroxides; or such organic bases as isopropylamine, trimethylamine, histidine or procaine.

The present invention contemplates vaccines for use in both active and passive immunization, in certain embodiments. Immunogenic compositions, proposed to be suitable for use as a vaccine, may be prepared most readily directly from attenuated virus of the *Flaviviridae* family with a mutant or altered viral genome, prepared in a manner disclosed herein. In various embodiments, an antigenic material may be extensively dialyzed to remove undesired small molecular weight molecules and/or lyophilized for more ready formulation into a desired vehicle.

Typically, vaccines are prepared as injectables. Either as liquid solutions or suspensions: solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified. In embodiments where the composition is in a liquid form, a carrier can be a solvent or dispersion medium comprising but not limited to, water, ethanol, polyol (e.g., glycerol, propylene glycol,

liquid polyethylene glycol, etc), lipids (e.g., triglycerides, vegetable oils, liposomes) and combinations thereof. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin; by the maintenance of the required particle size by dispersion in carriers such as, for example liquid polyol or lipids; by the use of surfactants such as, for example hydroxypropylcellulose; or combinations thereof such methods. In many cases, it will be preferable to include isotonic agents, such as, for example, sugars, sodium chloride or combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, or adjuvants which enhance the effectiveness of the vaccines. Additionally, iscom, a supramolecular spherical structure, may be used for parenteral and mucosal vaccination (Morein et al., 1998).

Sterile injectable solutions may be prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and/or the other ingredients. In the case of sterile powders for the preparation of sterile injectable solutions, suspensions or emulsion, the preferred methods of preparation are vacuum-drying or freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered liquid medium thereof. The liquid medium should be suitably buffered if necessary and the liquid diluent first rendered isotonic prior to injection with sufficient saline or glucose. The preparation of highly concentrated compositions for direct injection is also contemplated, where the use of DMSO as solvent is envisioned to result in extremely rapid penetration, delivering high concentrations of the active agents to a small area.

Various methods of achieving adjuvant effect for the vaccine includes use of agents such as aluminum hydroxide or phosphate (alum), commonly used as about 0.05 to about 0.1% solution in phosphate buffered saline, admixture with synthetic polymers of sugars (Carbopol®) used as an about 0.25% solution, aggregation of the protein in the vaccine by heat treatment with temperatures ranging between about 70° to about 101°C for a 30-second to 2-minute period, respectively. Aggregation by reactivating with

pepsin treated (Fab) antibodies to albumin, mixture with bacterial cells such as C. parvum or endotoxins or lipopolysaccharide components of gram-negative bacteria, emulsion in physiologically acceptable oil vehicles such as mannide mono-oleate (Aracel A) or emulsion with a 20% solution of a perfluorocarbon (Fluosol-DA®) used as a block substitute may also be employed.

Adjuvants that may be used in the practice of the invention include, but are not limited to Adjumer™, Adju-Phos, Algal Glucan, Algammulin, Alhydrogel, Antigen Formulation, Avridine®, BAY R1005,. Calcitriol, Calcium Phosphate Gel, Cholera holotoxin (CT), Cholera toxin B subunit (CTB), Cholera toxin A1-subunit-Protein A D-fragment fusion protein, CRL1005, Cytokine-containing Liposome, Dimethyldioctadecylammonium bromide, Dehydroepiandrosterone; Dimyristoyl phosphatidylcholine; 1,2-dimyristoyl-sn-3-phosphatidylcholine, Dimyristoyl phosphatidylglycerol, Deoxycholic Acid Sodium Salt; Freund's Complete Adjuvant, Freund's Incomplete Adjuvant, Gamma Inulin, Gerbu Adjuvant, GM-CSF, N-acetylglucosaminyl-(β1-4)-N-acetylmuramyl-L-alanyl-D-isoglutamine, Imiquimod, ImmTherTM, Interferon-γ, Interleukin-1β, Interleukin-2, Interleukin-7, Interleukin-12, ISCOMTM, Iscoprep 7.0.3.TM, Liposome, Loxoribine, LT-OA or LT Oral Adjuvant, MF59, MONTANIDE ISA 51, MONTANIDE ISA 720, MPLTM, MTP-PE, MTP-PE Liposome, Murametide, Murapalmitine, D-Murapalmitine, NAGO, Non-Ionic Surfactant Vesicle, Pleuran, lactic acid polymer,glycolic acid polymer, Pluronic L121, Polymethyl methacrylate, PODDSTM, Poly rA:Poly rU, Polysorbate 80, Protein Cochleate, QS-21, Quil-A, Rehydragel HPA, Rehydragel LV, S-28463, SAF-1, Sclavo peptide, Sendai Proteoliposome, Sendai-containing Lipid Matrix, Span 85, Specol, Squalane, Squalene, Stearyl Tyrosine, Theramide™, Threonyl-MDP, Ty Particle, or Walter Reed Liposome.

Any of the conventional methods for administration of a vaccine are applicable. These include, but are not limited to oral application on a solid physiologically acceptable base or in a physiologically acceptable dispersion, parenterally, by injection or the like. Vaccines of the invention may be administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include, in some cases, oral formulations. In other embodiments, one may use eye drops, nasal solutions or sprays, aerosols or inhalants in

the present invention. Such compositions are generally designed to be compatible with the target tissue type. In a non-limiting example, nasal solutions are usually aqueous solutions designed to be administered to the nasal passages in drops or sprays. Nasal solutions are prepared so that they are similar in many respects to nasal secretions, so that normal ciliary action is maintained. Thus, in preferred embodiments the aqueous nasal solutions usually are isotonic or slightly buffered to maintain a pH of about 5.5 to about 6.5. In addition, antimicrobial preservatives, similar to those used in ophthalmic preparations, drugs, or appropriate drug stabilizers, if required, may be included in the formulation. For example, various commercial nasal preparations are known and include drugs such as antibiotics or antihistamines.

In certain embodiments, the attenuated virus of the invention is prepared for administration by such routes as oral ingestion. In these embodiments, the solid composition may comprise, for example, solutions, suspensions, emulsions, tablets, pills, capsules (e.g., hard or soft shelled gelatin capsules), sustained release formulations, buccal compositions, troches, elixirs, suspensions, syrups, wafers, or combinations thereof. Oral compositions may be incorporated directly with the food of the diet. Preferred carriers for oral administration comprise inert diluents, assimilable edible carriers or combinations thereof. In other aspects of the invention, the oral composition may be prepared as a syrup or elixir. A syrup or elixir, and may comprise, for example, at least one active agent, a sweetening agent, a preservative, a flavoring agent, a dye, a preservative, or combinations thereof.

In certain preferred embodiments, an oral composition may comprise one or more binders, excipients, disintegration agents, lubricants, flavoring agents, and combinations thereof. In certain embodiments, a composition may comprise one or more of the following: a binder, such as, for example, gum tragacanth, acacia, cornstarch, gelatin or combinations thereof; an excipient, such as, for example, dicalcium phosphate, mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate or combinations thereof; a disintegrating agent, such as, for example, corn starch, potato starch, alginic acid or combinations thereof; a lubricant, such as, for example, magnesium stearate; a sweetening agent, such as, for example, sucrose, lactose, saccharin or combinations thereof; a flavoring agent, such as, for example peppermint, oil of

wintergreen, cherry flavoring, orange flavoring, etc.; or combinations thereof the foregoing. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, carriers such as a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar or both. Oral formulations may contain about 10 to about 95% of active ingredient, preferably about 25 to about 70%.

In certain embodiments, vaccines may comprise, for example, at least about 0.1% of an active compound. In other embodiments, an active compound may comprise between about 2% to about 75% of the weight of the unit, or between about 25% to about 60%, for example, and any range derivable therein. In other non-limiting examples, a dose may also comprise from about 1 microgram/kg/body weight, about 5 microgram/kg/body weight, about 10 microgram/kg/body weight, about 50 microgram/kg/body weight, about 100 microgram/kg/body weight, about 200 microgram/kg/body weight, about 350 microgram/kg/body weight, about 500 microgram/kg/body weight, about 1 milligram/kg/body weight, about 5 milligram/kg/body weight, about 10 milligram/kg/body weight, about 50 milligram/kg/body weight, about 100 milligram/kg/body weight, about 200 milligram/kg/body weight, about 350 milligram/kg/body weight, about 500 milligram/kg/body weight, to about 1000 mg/kg/body weight or more of antigen or total protein per administration, and any range derivable therein. In non-limiting examples of a derivable range from the numbers listed herein, a range of about 5 mg/kg/body weight to about 100 mg/kg/body weight, about 5 microgram/kg/body weight to about 500 milligram/kg/body weight, etc., can be administered, based on the numbers described above.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective and immunogenic. The quantity to be administered depends on the subject to be treated, including, e.g., the capacity of the individual's immune system to synthesize antibodies, and the degree of protection desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner. However, suitable dosage ranges are of the

order of several hundred micrograms active ingredient per vaccination. Suitable regimes for initial administration and booster shots are also variable, but are typified by an initial administration followed by subsequent inoculations or other administrations.

In many instances, it will be desirable to have multiple administrations of the vaccine, usually not exceeding six vaccinations, more usually not exceeding four vaccinations and preferably one or more, usually at least about three vaccinations. The vaccinations will normally be at from two to twelve week intervals, more usually from three to five week intervals. Periodic boosters at intervals of 1-5 years, usually three years, will be desirable to maintain protective levels of the antibodies. The course of the immunization may be followed by assays for antibodies for the supernatant antigens. The assays may be performed by labeling with conventional labels, such as radionuclides, enzymes, fluorescents, and the like. These techniques are well known and may be found in a wide variety of patents, such as U.S. Patent Nos. 3,791,932; 4,174,384 and 3,949,064, as illustrative of these types of assays.

"Unit dose" is defined as a discrete amount of a therapeutic composition dispersed in a suitable carrier. For example, in accordance with the present methods, viral doses include a particular number of viral or plaque forming units (pfu). For embodiments involving virus, particular unit doses include 10^1, $10^2 10^3$, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}, 10^{13}, 10^{14} or 10^{15} pfu or viral particles (vp).

In this connection, sterile aqueous media which can be employed will be known to those of skill in the art in light of the present disclosure. For example, a unit dose could be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. Moreover, for human administration, preparations should meet sterility, pyrogenicity, general safety and purity standards as required by FDA Office of Biologics standards.

The composition must be stable under the conditions of manufacture and storage, and preserved against the contaminating action of microorganisms, such as bacteria and

fungi. It will be appreciated that endotoxin contamination should be kept minimally at a safe level, for example, less that 0.5 ng/mg protein.

In particular embodiments, prolonged absorption of an injectable composition can be brought about by the use in the compositions of agents delaying absorption, such as, for example, aluminum monostearate, gelatin or combinations thereof.

IV. SCREENING ASSAYS

The present invention also contemplates the screening of compounds for various abilities to interact and/or affect *flavivirus*, in particular Yellow Fever virus, function and/or infectivity. Particularly preferred compounds will be those useful in inhibiting viral infection of cells, tissues, or organs. In the screening assays of the present invention, the candidate substance may first be screened for basic biochemical activity - *e.g.*, binding to Yellow Fever virus - and then tested for its ability to modulate activity or infectivity, at the cellular, tissue or whole animal level.

A. Assay Formats

The present invention provides methods of screening for modulators of yellow fever virus infectivity. In one embodiment, the present invention is directed to a method of:

(i) providing a Yellow Fever virus;
(ii) contacting the Yellow Fever virus with a candidate substance; and
(iii) determining the binding of the candidate substance to the Yellow Fever virus.

In yet another embodiment, the assay looks not at binding, but at viral infectivity. Such methods would comprise, for example:

(i) providing a cell that is susceptible to Yellow Fever virus infection;
(ii) contacting the virus with the candidate substance; and
(iii) determining the effect of the candidate substance on infectivity of Yellow Fever virus.

In still yet other embodiments, one would look at the effect of a candidate substance on the activity of Yellow Fever virus. This may involve looking at any of a number of characteristics, including Yellow Fever virus gene expression. An exemplary assay may include the detection of Yellow Fever virus nucleic acid by PCR.

B. Candidate Substances

As used herein, the term "candidate substance" refers to any molecule that may potentially modulate Yellow Fever virus infectivity. The candidate substance may be a protein or fragment thereof, a small molecule inhibitor, or even a nucleic acid molecule. It may prove to be the case that the most useful pharmacological compounds will be compounds that are structurally related to compounds which interact naturally with Yellow Fever virus or its family members. Creating and examining the action of such molecules is known as "rational drug design," and include making predictions relating to the structure of target molecules.

The goal of rational drug design is to produce structural analogs of biologically active polypeptides or target compounds. By creating such analogs, it is possible to fashion drugs which are more active or stable than the natural molecules, which have different susceptibility to alteration or which may affect the function of various other molecules. In one approach, one would generate a three-dimensional structure for a molecule like yellow virus envelope protein, and then design a molecule for its ability to interact with the envelope protein. Alternatively, one could design a partially functional fragment of an envelope protein (binding but no activity), thereby creating a competitive inhibitor. This could be accomplished by x-ray crystallography, computer modeling or by a combination of both approaches.

It also is possible to use antibodies to ascertain the structure of a target compound or inhibitor. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of anti-idiotype would be expected to be an analog of the original antigen. The anti-idiotype could then be used to identify

and isolate peptides from banks of chemically- or biologically-produced peptides. Selected peptides would then serve as the pharmacore. Anti-idiotypes may be generated using the methods described herein for producing antibodies, using an antibody as the antigen.

Candidate compounds may include fragments or parts of naturally-occurring compounds or may be found as active combinations of known compounds which are otherwise inactive. It is proposed that compounds isolated from natural sources, such as animals, bacteria, fungi, plant sources, including leaves and bark, and marine samples may be assayed as candidates for the presence of potentially useful pharmaceutical agents. It will be understood that the pharmaceutical agents to be screened could also be derived or synthesized from chemical compositions or man-made compounds. Thus, it is understood that the candidate substance identified by the present invention may be polypeptide, polynucleotide, small molecule inhibitors or any other compounds that may be designed through rational drug design starting from known inhibitors of a steroid hormone receptor repressor.

It will, of course, be understood that all the screening methods of the present invention are useful in themselves notwithstanding the fact that effective candidates may not be found. The invention provides methods for screening for such candidates, not solely methods of finding them.

EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLE 1:

METHODS

Animals:

The animals used in these studies were 3-4 week-old, female, Syrian golden hamsters (*Mesocricetus auratus*) from Harlan Sprague Dawley.

Hamster passages:

A single hamster was inoculated intraperitoneally (i.p.) with Asibi virus. At 3 days post infection (dpi), the liver was harvested and homogenized in PBS. After freezing at -70° overnight, 100µl of the liver homogenate was inoculated i.p. into a naïve hamster and is termed liver-to-liver passage. This process was repeated 6 times to generate the viscerotropic Asibi/hamster p7 virus.

Titration of viruses:

Serum was obtained by saphenous vein bleed each day for 6 days following i.p. inoculation with either wild-type Asibi/hamster p0 or viscerotropic Asibi/hamster p7. Virus titer in the serum was determined by tissue culture infectious dose 50% (TCID50) in Vero cells.

Morbidity and mortality:

Hamsters were inoculated i.p. with Asibi/hamster p0 or Asibi/hamster p7 virus and observed for signs of illness for 14 days. Signs of illness included: ruffled fur, lethargy, hunched posture, and paralysis. Some animals found to be completely moribund were euthanized to collect organ samples. These animals are not included in the survival curve.

Histopathology:

Liver and spleen were harvested 5 and 6 dpi for histological examination. The tissues were fixed in 10% buffered formalin for 48 hours and then transferred to 70% ethanol for storage. The tissues were paraffin embedded, sectioned, and stained with hematoxylin and eosin by the core facility (UTMB).

Sequence analysis:

Viral RNA was isolated using the QIamp viral RNA mini kit (Qiagen). The genome was amplified by RT-PCR with YF virus specific primers. Fragments were

cloned into either pGEM-T (Promega) or pCR (Invitrogen) vector and amplified in DH5α competent cells. A consensus sequence was taken from 3 or more clones sequenced in both directions. Automated sequencing was performed in the UTMB core laboratory. Sequence analysis was performed using the Vector NTI program (InforMax).

EXAMPLE 2

Passage of wild-type YF virus Asibi in hamsters:

Wild-type, non-hamster passaged Asibi (p0) virus causes a mild and transient viremia with no outward signs of illness in sub-adult hamsters. The 7th hamster passage (Asibi/hamster p7) virus was found to be highly viscerotropic in hamsters and caused severe illness and death in 100% of sub-adult hamsters.

Morbidity and mortality:

Sub-adult hamsters were inoculated with either the parental Asibi/hamster p0 or the hamster-viscerotropic Asibi/hamster p7 virus and observed for 14 days. Hamsters inoculated with Asibi/hamster p0 virus showed no outward signs of illness, and all animals survived. In contrast all 7 hamsters inoculated with Asibi/hamster p7 developed outward signs of illness including ruffled fur, lethargy, and hunched posture and died within 2 days of onset of clinical signs of disease. Signs of illness appeared as early as 2 dpi, and all animals succumbed to illness by 8 dpi. The survival of these animals is summarized in FIG. 1.

Viremia:

Hamsters inoculated with Asibi/hamster p7 virus developed a robust viremia that peaked at 3dpi (FIG. 2), as shown with other strains of YF virus by Tesh et al. (2001). Only a modest viremia was detected in hamsters inoculated with Asibi/hamster p0, and no viremia was detected in 2 of 5 animals (FIG. 2).

Histopathology

Spleen and liver were harvested on 5-6 dpi (at a time determined by Xiao et al (2001) to be the peak of histopathologic changes). Samples from 5 animals (A-E) inoculated with either Asibi/hamster p0 or Asibi/hamster p7 were paraffin embedded and stained with hematoxylin and eosin for microscopic evaluation.

Liver-

The livers of hamsters inoculated with Asibi/hamster p0 showed no significant changes on either day 5 or 6 pi when compared with mock-infected animals (FIG. 3A and 3B). However, the livers of hamsters inoculated with Asibi/hamster p7 showed significant pathologic changes including microvesicular steatosis, moderate to severe inflammation, and mild to moderate hepatic necrosis (FIG. 3C). Hamster A was sacrificed on day 5 pi due to extreme illness. The remaining four hamsters were sacrificed on day 6 pi when they were beginning to show clinical signs of illness. The liver of hamster A had the most pronounced steatosis (98%) and severe hepatic necrosis with only mild/moderate inflammation. The steatosis in the other hamsters involved 50-95% of the liver, with an average of 77% (FIG. 4A). The results of these studies are summarized in FIG. 4.

Spleen-

The spleens of 4 out of 5 hamsters inoculated with Asibi/hamster p0 were characterized by marked lymphoid hyperplasia and moderate to severe white pulp depletion, necrosis, and splenic macrophage hyperplasia (FIG. 5 and 6). The spleen from hamster E showed no abnormalities. There was no lymphoid hyperplasia in any of the spleens from hamsters inoculated with Asibi/hamster p7; however, there was severe splenic macrophage hyperplasia and necrosis. There was also moderate to severe white pulp depletion (FIG. 5 and 6).

EXAMPLE 3

Nucleotide and deduced amino acid changes of Asibi/hamster p7 virus

The complete genomic sequence of the Asibi/hamster p7 virus was determined and compared with that of the published Asibi sequence (Hahn *et al.*, 1987) identifying 23 nucleotide changes. Regions that contained nucleotide changes were amplified from the parental Asibi/hamster p0 virus and sequenced for confirmation. The parental Asibi strain used in these studies was obtained from the World Reference Center. This virus differed from the published sequence for Asibi (Hahn *et al.*, 1987) at genomic nucleotide positions 2193, 2355, 2704, 3817, 3925, 5926, 6013, 6829, and 7319 reducing the

number of nucleotide changes between the parental and hamster-passaged viruses from 23 to 14. These 14 nucleotide changes encoded 7 amino acid substitutions (Table 2).

Table 2: Summary of the nucleotide and deduced amino acid changes between Asibi/hamster p0 virus and Asibi/hamster p7 virus.

Nucleotide	Asibi p0	Asibi p7	amino acid	Asibi p0	Asibi p7
802	A	G			
887[b]	C	U			
1000[b]	G	A			
1054[b]	A	C	E27	Q	H
1056	A	G	E28	D	G
1437[b]	A	C	E155	D	A
1941	A	G	E323	K	R
1965[a,b]	A	G	E331	K	R
2779	U	C			
3274[a,b]	G	A			
3821[b,c]	A	G	NS2A48	T	A
4864[a,b]	G	A			
7178[b,c]	G	A	NS4B98	V	I
8917	C	U			

[a] Nucleotide substitutions shared with 17D virus
[b] Nucleotide substitutions shared with Asibi/HeLa p6 virus
[c] Nucleotide substitutions shared with FNV virus

The nucleotide and amino acid substitutions in the Asibi/hamster p7 virus were not evenly distributed throughout the genome (Table 3). No nucleotide changes were identified in the 5' or 3' NCR of the Asibi/hamster p7 virus. There were 8 nucleotide substitutions found within the structural protein genes (2 in the M protein gene and 6 in the E protein gene), and the remaining 6 nucleotide changes were located within the non-structural protein genes (2 in NS1; 1 each in NS2A, NS3, NS4B and NS5). No nucleotide changes were identified within the C, prM, NS2B, NS4A, and 2K protein genes or within the 5' or 3' non-coding regions (NCR). Two amino acid substitutions were located in the non-structural proteins at positions NS2A48 (T to A), and NS4B98 (V to I); however, the majority of the amino acid changes were located in the E protein: E27 (Q to H), E28 (D to G), E155 (D to A), E323 (K to R), E331 (K to R). Only certain

regions of the genome can tolerate mutation; therefore, viable viruses accumulate mutations only within these regions despite strong selective pressures. Many of the nucleotide and amino acid changes identified in the Asibi/hamster p7 virus are common to other derivatives of Asibi (17D and Asibi/HeLa p6), the vaccine strain FNV, and/or wild-type YF viruses. Only 4 nucleotide changes appear to be unique to the Asibi/hamster p7 virus (Table 2) and these encode 1 amino acid substitution at E323 (K to R). A genbank search for amino acids common to those found in Asibi/hamster p7 virus revealed only 3 YF isolates from the East and Central African genotype with a glycine residue at position E28 (Ethiopia 60A and 60B, and CAR 80) (Mutebi *et al.*, 2001). All YF virus sequences in genbank had lysine at residue E 323 where Asibi/hamster p7 had an arginine residue.

Table 3: Distribution of nucleotide and amino acid changes throughout the genome of the Asibi/hamster p7 virus

Region	Length	total ntd changes	% ntd changes	Total aa changes	% aa changes
5'NCR	119	0	*	0	*
C	362	0	*	0	*
PrM	267	0	*	0	*
M	225	2	0.8	0	*
E	1479	6	0.4	5	1.0
NS1	1227	2	0.2	0	*
NS2A	501	1	0.2	1	0.6
NS2B	390	0	*	0	*
NS3	1869	1	0.1	0	*
NS4A	378	0	*	0	*
2K	66	0	*	0	*
NS4B	750	1	0.1	1	0.4
NS5	2715	1	0.04	0	*
3'NCR	511	0	*	0	*

Seven of the 14 nucleotide changes encode amino acid substitutions, and 5 of these are located in the E protein at amino acid positions: E27 (Q to H), E28 (D to G), E155 (D to A), E323 (K to R), E331 (K to R). The location of these amino acid changes has been modeled onto the TBE virus E protein crystal structure (FIG. 7) to investigate the potential interactions of the amino acid substitutions. E27, E28, and E155 are located

25176161.1

in domain I with E27 and E28 adjacent to one another and E155 spatially distinct. The other 2 changes E323 and E331 are located relatively close together in domain III. There are no amino acid substitutions within domain II or the stem-anchor region.

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

U. S. Patent 3,791,932

U. S. Patent 3,949,064

U. S. Patent 4,174,384

U. S. Patent 4,500,512

U. S. Patent 4,554,101

U. S. Patent 4,683,195

U. S. Patent 4,683,202

U. S. Patent 4,800,159

U. S. Patent 4,810,492

U. S. Patent 4,883,750

U. S. Patent 4,946,773

U. S. Patent 5,220,007

U. S. Patent 5,279,721

U. S. Patent 5,284,760

U. S. Patent 5,354,670

U. S. Patent 5,366,878

U. S. Patent 5,389,514

U. S. Patent 5,635,377

U. S. Patent 5,736,148

U. S. Patent 5,744,140

U. S. Patent 5,744,141

U. S. Patent 5,789,166

U. S. Patent 5,798,208

U. S. Patent 5,830,650

U. S. Patent 5,840,873

U. S. Patent 5,843,640
U. S. Patent 5,843,650
U. S. Patent 5,843,651
U. S. Patent 5,843,663
U. S. Patent 5,846,708
U. S. Patent 5,846,709
U. S. Patent 5,846,717
U. S. Patent 5,846,726
U. S. Patent 5,846,729
U. S. Patent 5,846,783
U. S. Patent 5,849,481
U. S. Patent 5,849,483
U. S. Patent 5,849,486
U. S. Patent 5,849,487
U. S. Patent 5,849,497
U. S. Patent 5,849,546
U. S. Patent 5,849,547
U. S. Patent 5,851,770
U. S. Patent 5,851,772
U. S. Patent 5,853,990
U. S. Patent 5,853,992
U. S. Patent 5,853,993
U. S. Patent 5,856,092
U. S. Patent 5,858,652
U. S. Patent 5,861,244
U. S. Patent 5,863,732
U. S. Patent 5,863,753
U. S. Patent 5,866,331
U. S. Patent 5,866,337
U. S. Patent 5,866,366
U. S. Patent 5,900,481

U. S. Patent 5,905,024
U. S. Patent 5,910,407
U. S. Patent 5,912,124
U. S. Patent 5,912,145
U. S. Patent 5,912,148
U. S. Patent 5,916,776
U. S. Patent 5,916,779
U. S. Patent 5,919,626
U. S. Patent 5,919,630
U. S. Patent 5,922,574
U. S. Patent 5,925,517
U. S. Patent 5,925,525
U. S. Patent 5,928,862
U. S. Patent 5,928,869
U. S. Patent 5,928,870
U. S. Patent 5,928,905
U. S. Patent 5,928,906
U. S. Patent 5,929,227
U. S. Patent 5,932,413
U. S. Patent 5,932,451
U. S. Patent 5,935,791
U. S. Patent 5,935,825
U. S. Patent 5,939,291
U. S. Patent 5,942,391
U. S. Patent 6,171,854
U. S. Patent 6,184,024
U. S. Patent 6,254,873
U. S. Patent 6,337,073
U. S. Patent 6,372,221
Bellus, *J. Macromol. Sci. Pure Appl. Chem.*, A31(1):1355-1376, 1994.
Brown *et al.*, *J. Immunol.*, 156(9):3285-3291, 1996.

Burton and Barbas, *Adv. Immunol.*, 57:191-280, 1994.

Cooley *et al., Science*, 239(4844):1121-1128, 1988.

Cunningham and Wells, *Science*, 244(4908):1081-1085, 1989.

European Appl. 320 308

European Appl. 329 822

Frohman, In: *PCR Protocols: A Guide To Methods And Applications,* Academic Press, NY, 1990.

GB Appl. 2, 202 328

Hahn *et al., Proc. Natl. Acad. Sci. USA*, 84:2019-2023, 1987.

Harlow and Lane, In: *Antibodies: A Laboratory Manuel,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988.

Hilton *et al., J. Biol. Chem.*, 271(9):4699-4708, 1996.

Innis *et al., Proc. Natl. Acad. Sci. USA*, 85(24):9436-9440, 1988.

Jackson *et al., Plant J.*, 8(3):369-380, 1995.

Koncz *et al., EMBO J.*, 9(5):1337-1346, 1990.

Kwoh *et al., Proc. Nat. Acad. Sci. USA*, 86: 1173, 1989.

Kyte and Doolittle, *J. Mol. Biol.*, 157(1):105-132, 1982.

Marks *et al., Symp. Soc. Exp. Biol.*, 45:77-87, 1991.

Morein *et al., Dev. Biol. Stand.*, 92:33-39, 1998.

Mutebi *et al., J. Virol.*, 75:6999-7008, 2001.

Ohara *et al., Proc. Natl. Acad. Sci. USA*, 86:5673-5677, 1989.

Oppenheimer *et al., Cell*, 67(3):483-493, 1991.

PCT Appl. PCT/US87/00880

PCT Appl. PCT/US89/01025

PCT Appl. WO 88/10315

PCT Appl. WO 89/06700

Remington's Pharmaceutical Sciences, 15th ed., 1035-1038 and 1570-1580, Mack Publishing Company, Easton, PA, 1980.

Remington's Pharmaceutical Sciences, 18th Ed. Mack Printing Company, 1289-1329, 1990.

Rey *et al., Nature*, 375:291-298, 1995.

Sambrook *et al.*, *In: Molecular cloning,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Schmidt *et al., Science*, 238(4829):960-963, 1987.

Sommer *et al., EMBO J.*, 9(3):605-13, 1990.

Tesh *et al., J. Infect. Dis.* 183:1431-1436, 2001.

Walker *et al., Proc. Natl. Acad. Sci. USA,* 89:392-396 1992.

Warren *et al., Biochemistry*, 35(27):8855-8862, 1996.

Wong *et al., J. Bacteriol.*, 178(8):2334-2342, 1996.

Xiao *et al., Pathol. J. Infect. Dis.*, 183:1437-1444, 2001.

Yelton *et al., J. Immunol.*, 155(4):1994-2004, 1995.

Zeng *et al., Biochemistry*, 35(40):13157-13164, 1996.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid encoding a Yellow Fever virus with a viral genome that comprises at least one of the following alterations:

   a) an alteration in the nucleic acid sequence resulting in an envelope protein with a histidine at amino acid 27;

   b) an alteration in the nucleic acid sequence resulting in an envelope protein with a glycine at amino acid 28;

   c) an alteration in the nucleic acid sequence resulting in an envelope protein with a alanine at amino acid 155;

   d) an alteration in the nucleic acid sequence resulting in an envelope protein with an arginine at amino acid 323;

   e) an alteration in the nucleic acid sequence resulting in an envelope protein with an arginine at amino acid 331;

   f) an alteration in the nucleic acid sequence resulting in a NS2A protein with an alanine at amino acid 48; or

   g) an alteration in the nucleic acid sequence resulting in a NS4B protein with an isoleucine at amino acid 98.

2. The nucleic acid of claim 1, wherein the nucleic acid is RNA.

3. The nucleic acid of claim 1, wherein the nucleic acid is DNA.

4. The nucleic acid of claim 1, wherein the viral genome comprises at least two of alterations a-g.

5. The nucleic acid of claim 1, wherein the viral genome comprises at least three of alterations a-g.

6. The nucleic acid of claim 1, wherein the viral genome comprises at least four of alterations a-g.

7. The nucleic acid of claim 1, wherein the viral genome comprises at least five of alterations a-g.

8. The nucleic acid of claim 1, wherein the viral genome comprises at least six of alterations a-g.

9. The nucleic acid of claim 1, wherein the viral genome comprises seven of alterations a-g.

10. The nucleic acid of claim 1, wherein the nucleic acid has a nucleic acid sequence as set forth in SEQ ID NO:1.

11. A isolated nucleic acid comprising 10 to 200 contiguous nucleotides of SEQ ID NO:1.

12. The isolated nucleic acid of claim 11, wherein said nucleic acid comprises 15 to 150 contiguous nucleotides.

13. The isolated nucleic acid of claim 11, wherein said nucleic acid comprises 20 to 100 contiguous nucleotides.

14. The isolated nucleic acid of claim 11, wherein said nucleic acid comprises 25 to 50 contiguous nucleotides.

15. A vaccine composition comprising a Yellow Fever virus with a viral genome that comprises at least one of the following alterations:

a) an alteration in a nucleic acid sequence encoding amino acid 323 of an/the envelope protein, wherein the first alteration requires more than one nucleotide change to encode an arginine;

b) an alteration in a nucleic acid sequence encoding amino acid 27 of an/the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a histidine;

c) an alteration in a nucleic acid sequence encoding amino acid 28 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a glycine;

d) an alteration in a nucleic acid sequence encoding amino acid 155 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode an alanine;

e) an alteration in a nucleic acid sequence encoding amino acid 331 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode an arginine;

f) an alteration in a nucleic acid sequence encoding amino acid 48 of the NS2A protein, wherein the second alteration requires more than one nucleotide change to encode an alanine; or

g) an alteration in a nucleic acid sequence encoding amino acid 98 of the NS4B protein, wherein the second alteration requires more than one nucleotide change to encode an isoleucine.

16. The vaccine composition of claim 15, wherein the viral genome comprises at least two of alterations a-g.

17. The vaccine composition of claim 15, wherein the viral genome comprises at least three of alterations a-g.

18. The vaccine composition of claim 15, wherein the viral genome comprises at least four of alterations a-g.

19. The vaccine composition of claim 15, wherein the viral genome comprises at least five of alterations a-g.

20. The vaccine composition of claim 15, wherein the viral genome comprises at least six of alterations a-g.

21. The vaccine composition of claim 15, wherein the viral genome comprises seven of alterations a-g.

22. The vaccine composition of claim 15, wherein the composition is a pharmaceutically acceptable formulation.

23. The vaccine composition of claim 15, wherein the Yellow Fever virus is a 17D virus.

24. The vaccine composition of claim 15, wherein the Yellow Fever virus is a 17D-204 virus.

25. The vaccine composition of claim 15, wherein the Yellow Fever virus is a 17DD virus.

26. A method for producing an attenuated Yellow Fever virus comprising introducing into a Yellow Fever virus genome a missense mutation that would require two nucleotide changes to encode a supervirulence amino acid.

27. A method for producing a Yellow Fever virus vaccine comprising:

a) identifying a mutation that results in a missense mutation in a first Yellow Fever viral genome that is associated with an increased virulence of the virus;

b) modifying an attenuated Yellow Fever viral genome by mutation of a codon associated with the missense mutation resulting in a reduced probability of reversion to a virulent phenotype.

28. The method of claim 27, wherein the missense mutation results in an envelope protein having an arginine at amino acid position 323.

29. The method of claim 27, wherein modifying the attenuated Yellow Fever virus is by substituting a second codon that encodes for a conservative amino acid change.

30. A method for identifying a compound active against a viral infection comprising:

a) providing a virus expressed from a viral construct comprising a nucleic acid encoding a Yellow Fever virus comprising an envelope protein comprising an arginine at amino acid 323;

b) contacting the virus with a candidate substance; and

c) comparing the infectious ability of the virus in the presence of said candidate substance with the infectious ability of the virus in a similar system in the absence of the candidate substance.

31. The method of claim 30, wherein the nucleic acid encodes a virus with an envelope protein further comprising a histidine at amino acid 27, a glycine at amino acid 28, an alanine at amino acid 155, and an arginine at amino acid 331.

32. The method of claim 30, wherein the nucleic acid sequence is that set forth in SEQ ID NO:1.

33. A method of vaccination against a virus comprising administering to a subject a Yellow Fever virus with a viral genome that comprises at least one of the following alterations:

a) an alteration in the nucleic acid sequence encoding amino acid 323 of an/the envelope protein, wherein the first alteration requires more than one nucleotide change to encode an arginine;

b) an alteration in the nucleic acid sequence encoding amino acid 27 of an/the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a histidine;

c) an alteration in the nucleic acid sequence encoding amino acid 28 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode a glycine;

d) an alteration in the nucleic acid sequence encoding amino acid 155 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode an alanine;

e) an alteration in the nucleic acid sequence encoding amino acid 331 of the envelope protein, wherein the second alteration requires more than one nucleotide change to encode an arginine;

f) an alteration in the nucleic acid sequence encoding amino acid 48 of the NS2A protein, wherein the second alteration requires more than one nucleotide change to encode an alanine; or

g) an alteration in the nucleic acid sequence encoding amino acid 98 of the NS4B protein, wherein the second alteration requires more than one nucleotide change to encode an isoleucine.

34. The method of vaccination of claim 33, wherein the viral genome comprises at least two alterations.

35. The method of vaccination of claim 33, wherein the viral genome comprises at least three alterations.

36. The method of vaccination of claim 33, wherein the viral genome comprises at least four alterations.

37. The method of vaccination of claim 33, wherein the viral genome comprises at least five alterations.

38. The method of vaccination of claim 33, wherein the viral genome comprises at least six alterations.

39. The method of vaccination of claim 33, wherein the viral genome comprises seven alterations.

40. The method of vaccination of claim 33, wherein the composition is a pharmaceutically acceptable formulation.

41. The method of vaccination of claim 33, wherein the Yellow Fever virus is a 17D virus.

42. The method of vaccination of claim 33, wherein the Yellow Fever virus is a 17D-204 virus.

43. The method of vaccination of claim 33, wherein the Yellow Fever virus is a 17DD virus.

ABSTRACT

The present invention concerns the use of methods and/or compositions for the improvement of the reversion frequency of an attenuated member of the *Flaviviridae* family. In particular embodiments of the invention, methods and compositions of the invention may be used for the improvement and/or production of a Yellow Fever virus vaccine.

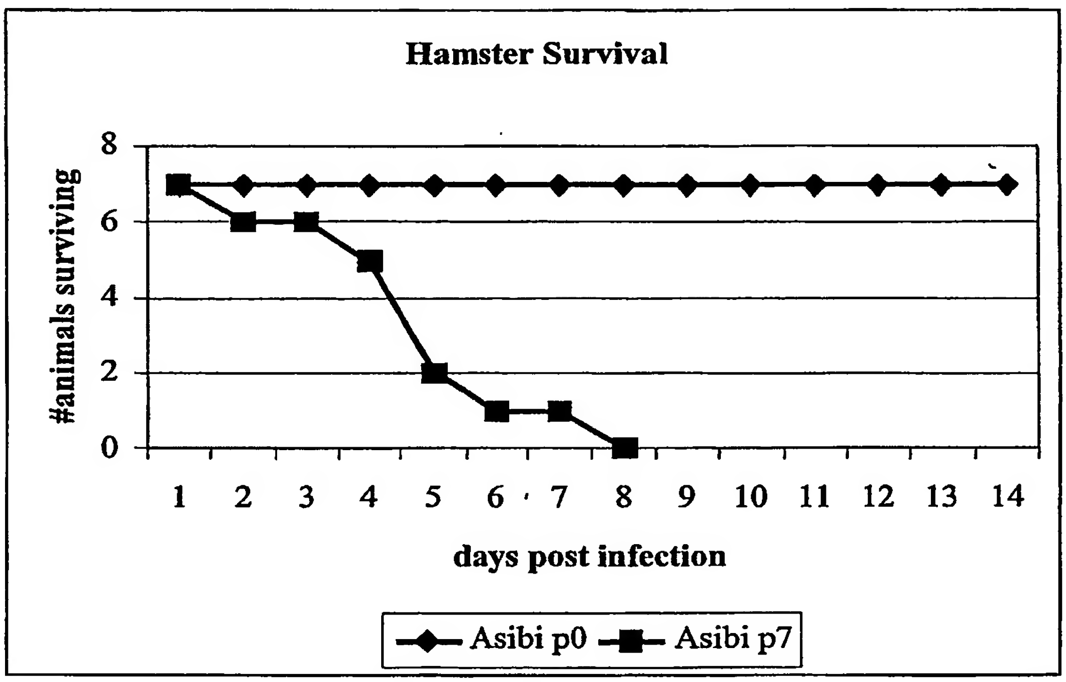

Hamster Survival
#animals surviving
0
2
4
6
8
1 2 3 4 5 6 7 8 9 10 11 12 13 14
days post infection
Asibi p0
Asibi p7

FIG. 1

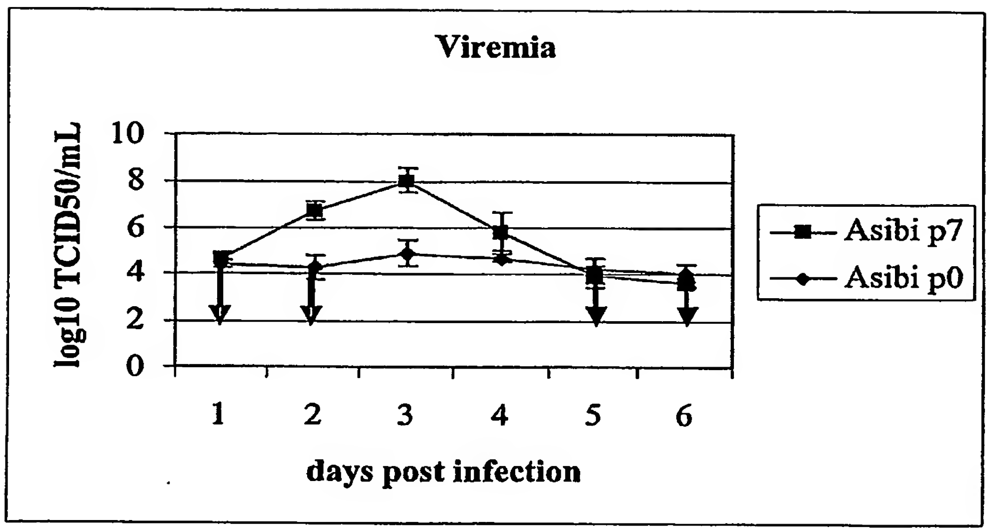

Viremia
log10 TCID50/mL
10
8
6
4
2
0
1
2
3
4
5
6
days post infection
Asibi p7
Asibi p0

FIG. 2

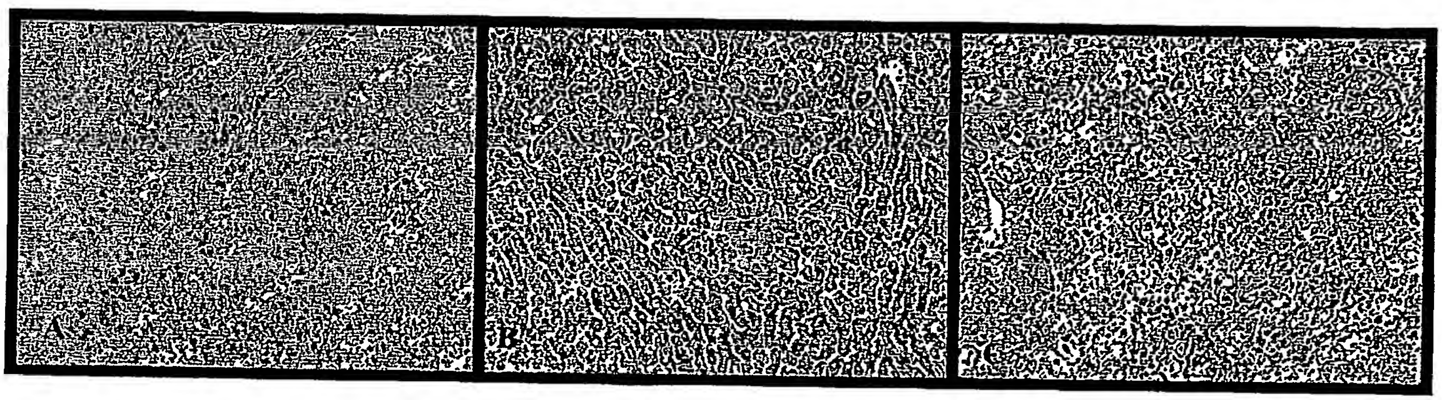
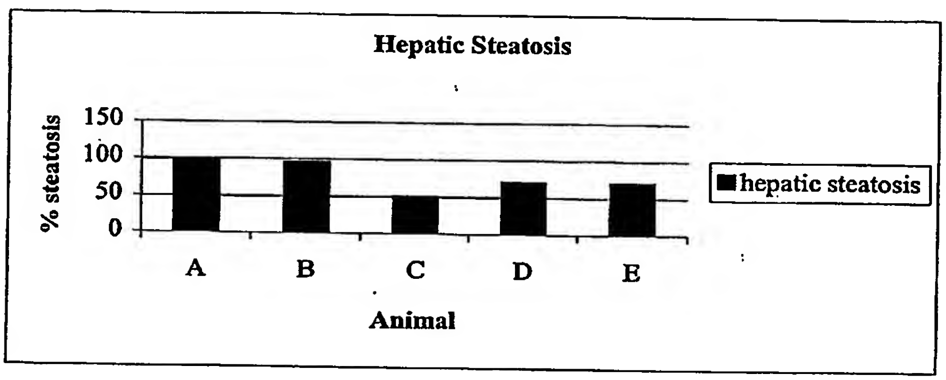


FIG. 3

A


Hepatic Steatosis
% steatosis
150
100
50
0
A
B
C
D
E
Animal
hepatic steatosis

B

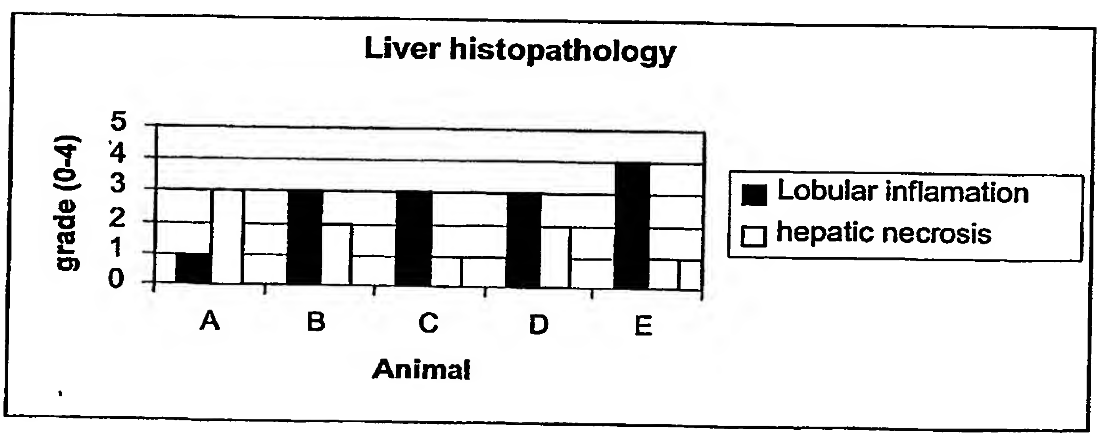

Liver histopathology
grade (0-4)
5
4
3
2
1
0
A
B
C
D
E
Animal
Lobular inflamation
hepatic necrosis

FIG. 4A-B

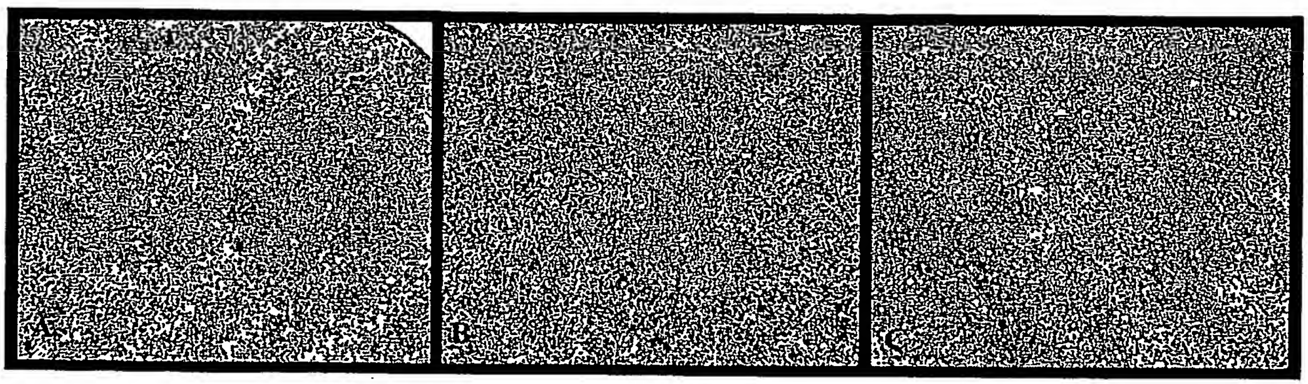
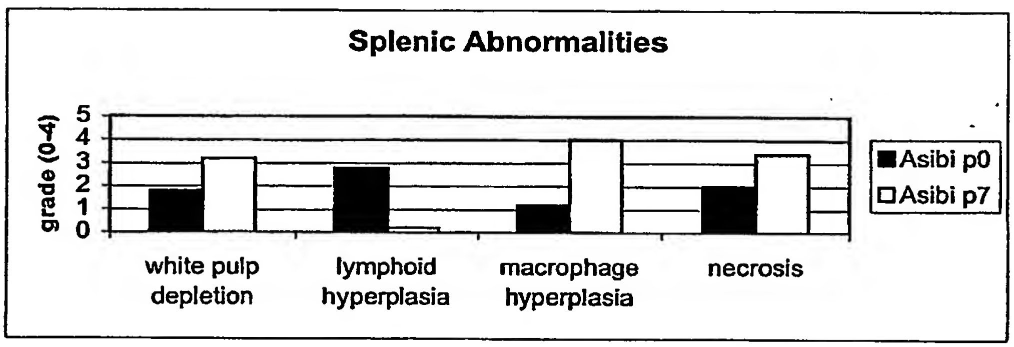


FIG. 5



Splenic Abnormalities
grade (0-4)
5
4
3
2
1
0
white pulp depletion
lymphoid hyperplasia
macrophage hyperplasia
necrosis
Asibi p0
Asibi p7

FIG. 6

FIG. 7

SEQUENCE LISTING

```
<110> BARRETT, ALAN
      MCARTHUR, MONICA

<120> METHODS AND COMPOSITIONS CONCERNING ALTERED YELLOW
      FEVER VIRUS STRAINS

<130> UTSG:255USP1

<140> UNKNOWN
<141> 2002-07-19

<160> 4

<170> PatentIn Ver. 2.1

<210> 1
<211> 10862
<212> DNA
<213> Yellow fever virus

<220>
<221> CDS
<222> (119)..(10744)

<400> 1
agtaaatcct gtgtgctaat tgaggtgcat tggtctgcaa atcgagttgc taggcaataa 60

acacatttgg attaatttta atcgttcgtt gagcgattag cagagaactg accagaac   118

atg tct ggt cgt aaa gct cag gga aaa acc ctg ggc gtc aat atg gta   166
Met Ser Gly Arg Lys Ala Gln Gly Lys Thr Leu Gly Val Asn Met Val
  1               5                  10                  15

cga cga gga gtt cgc tcc ttg tca aac aaa ata aaa caa aaa aca aaa   214
Arg Arg Gly Val Arg Ser Leu Ser Asn Lys Ile Lys Gln Lys Thr Lys
             20                  25                  30

caa att gga aac aga cct gga cct tca aga ggt gtt caa gga ttt atc   262
Gln Ile Gly Asn Arg Pro Gly Pro Ser Arg Gly Val Gln Gly Phe Ile
         35                  40                  45

ttt ttc ttt ttg ttc aac att ttg act gga aaa aag atc acg gcc cac   310
Phe Phe Phe Leu Phe Asn Ile Leu Thr Gly Lys Lys Ile Thr Ala His
     50                  55                  60

cta aag agg ttg tgg aaa atg ctg gac cca aga caa ggc ttg gct gtt   358
Leu Lys Arg Leu Trp Lys Met Leu Asp Pro Arg Gln Gly Leu Ala Val
 65                  70                  75                  80

cta agg aaa gtt aag aga gtg gtg gcc agt ttg atg aga gga ttg tcc   406
```

Leu Arg Lys Val Lys Arg Val Val Ala Ser Leu Met Arg Gly Leu Ser
85 90 95

tca agg aaa cgc cgt tcc cat gat gtt ctg act gtg caa ttc cta att 454
Ser Arg Lys Arg Arg Ser His Asp Val Leu Thr Val Gln Phe Leu Ile
100 105 110

ttg gga atg ctg ttg atg acg ggt gga gtg acc ttg gtg cgg aaa aac 502
Leu Gly Met Leu Leu Met Thr Gly Gly Val Thr Leu Val Arg Lys Asn
115 120 125

aga tgg ttg ctc cta aat gtg aca tct gag gac ctc ggg aaa aca ttc 550
Arg Trp Leu Leu Leu Asn Val Thr Ser Glu Asp Leu Gly Lys Thr Phe
130 135 140

tct gtg ggc aca ggc aac tgc aca aca aac att ttg gaa gcc aag tac 598
Ser Val Gly Thr Gly Asn Cys Thr Thr Asn Ile Leu Glu Ala Lys Tyr
145 150 155 160

tgg tgc cca gac tca atg gaa tac aac tgt ccc aat ctc agt cca aga 646
Trp Cys Pro Asp Ser Met Glu Tyr Asn Cys Pro Asn Leu Ser Pro Arg
165 170 175

gag gag cca gat gac att gat tgc tgg tgc tat ggg gtg gaa aac gtt 694
Glu Glu Pro Asp Asp Ile Asp Cys Trp Cys Tyr Gly Val Glu Asn Val
180 185 190

aga gtc gca tat ggt aag tgt gac tca gca ggc agg tct agg agg tca 742
Arg Val Ala Tyr Gly Lys Cys Asp Ser Ala Gly Arg Ser Arg Arg Ser
195 200 205

aga agg gcc att gac ttg cct acg cat gaa aac cat ggt ttg aag acc 790
Arg Arg Ala Ile Asp Leu Pro Thr His Glu Asn His Gly Leu Lys Thr
210 215 220

cgg caa gaa aag tgg atg act gga aga atg ggt gaa agg caa ctc caa 838
Arg Gln Glu Lys Trp Met Thr Gly Arg Met Gly Glu Arg Gln Leu Gln
225 230 235 240

aag att gag aga tgg ctc gtg agg aac ccc ttt ttt gca gtg aca gct 886
Lys Ile Glu Arg Trp Leu Val Arg Asn Pro Phe Phe Ala Val Thr Ala
245 250 255

ttg acc att gcc tac ctt gtg gga agc aac atg acg caa cga gtc gtg 934
Leu Thr Ile Ala Tyr Leu Val Gly Ser Asn Met Thr Gln Arg Val Val
260 265 270

att gcc cta ctg gtc ttg gct gtt ggt ccg gcc tac tca gct cac tgc 982
Ile Ala Leu Leu Val Leu Ala Val Gly Pro Ala Tyr Ser Ala His Cys
275 280 285

att gga att act gac aga gat ttc att gag ggg gtg cat gga gga act 1030

Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His Gly Gly Thr
290 295 300

tgg gtt tca gct acc ctg gag cac ggc aag tgt gtc act gtt atg gcc 1078
Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr Val Met Ala
305 310 315 320

cct gac aag cct tca ttg gac atc tca cta gag aca gta gcc att gat 1126
Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val Ala Ile Asp
325 330 335

gga cct gct gag gcg agg aaa gtg tgt tac aat gca gtt ctc act cat 1174
Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val Leu Thr His
340 345 350

gtg aag att aat gac aag tgc ccc agc act gga gag gcc cac cta gct 1222
Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala His Leu Ala
355 360 365

gaa gag aac gaa ggg gac aat gcg tgc aag cgc act tat tct gat aga 1270
Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr Ser Asp Arg
370 375 380

ggc tgg ggc aat ggc tgt ggc cta ttt ggg aaa ggg agc att gtg gca 1318
Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser Ile Val Ala
385 390 395 400

tgc gcc aaa ttc act tgt gcc aaa tcc atg agt ttg ttt gag gtt gat 1366
Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe Glu Val Asp
405 410 415

cag acc aaa att cag tat gtc atc aga gca caa ttg cat gta ggg gcc 1414
Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His Val Gly Ala
420 425 430

aag cag gaa aat tgg aat acc gcc att aag act ctc aag ttt gat gcc 1462
Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys Phe Asp Ala
435 440 445

ctg tca ggc tcc cag gaa gcc gag ttc act ggg tat gga aaa gct aca 1510
Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly Lys Ala Thr
450 455 460

ctg gaa tgc cag gtg caa act gcg gtg gac ttt ggt aac agt tac atc 1558
Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn Ser Tyr Ile
465 470 475 480

gct gag atg gaa aaa gag agc tgg ata gtg gac aga cag tgg gcc cag 1606
Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln Trp Ala Gln
485 490 495

gac ttg acc ctg cca tgg cag agt gga agt ggc ggg gtg tgg aga gag 1654

Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val Trp Arg Glu
500 505 510

atg cat cat ctt gtc gaa ttt gaa cct ccg cat gcc gcc act atc aga 1702
Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala Thr Ile Arg
515 520 525

gta ctg gcc ctg gga aac cag gaa ggc tcc ttg aaa aca gct ctt acc 1750
Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr Ala Leu Thr
530 535 540

ggc gca atg agg gtt aca aag gac aca aat gac aac aac ctt tac aaa 1798
Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn Leu Tyr Lys
545 550 555 560

cta cat ggt gga cat gtt tcc tgc aga gtg aaa ttg tca gct ttg aca 1846
Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser Ala Leu Thr
565 570 575

ctc aag ggg aca tcc tac aaa atg tgc act gac aaa atg tct ttt gtc 1894
Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met Ser Phe Val
580 585 590

aag aac cca act gac act ggc cat ggc act gtt gtg atg cag gtg aga 1942
Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met Gln Val Arg
595 600 605

gtg cca aaa gga gcc ccc tgc agg att cca gtg ata gta gct gat gat 1990
Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val Ala Asp Asp
610 615 620

ctt aca gcg gca atc aat aaa ggc att ttg gtt aca gtt aac ccc atc 2038
Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val Asn Pro Ile
625 630 635 640

gcc tca acc aat gat gat gaa gtg ctg att gag gtg aac cca cct ttt 2086
Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn Pro Pro Phe
645 650 655

gga gac agc tac att atc gtt ggg aca gga gat tca cgt ctc act tac 2134
Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg Leu Thr Tyr
660 665 670

cag tgg cac aaa gag gga agc tca ata gga aag ttg ttc act cag acc 2182
Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe Thr Gln Thr
675 680 685

atg aaa ggc gcg gaa cgc ctg gcc gtc atg gga gac gcc gcc tgg gat 2230
Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala Ala Trp Asp
690 695 700

ttc agc tcc gct gga ggg ttc ttc act tcg gtt ggg aaa gga att cat 2278

Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys Gly Ile His
705 710 715 720

acg gtg ttt ggc tct gcc ttt cag ggg cta ttt ggc ggc ttg aac tgg 2326
Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly Leu Asn Trp
725 730 735

ata aca aag gtc atc atg ggg gcg gta ctc ata tgg gtt ggc atc aac 2374
Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val Gly Ile Asn
740 745 750

aca aga aac atg aca atg tcc atg agc atg atc ttg gta gga gtg atc 2422
Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val Gly Val Ile
755 760 765

atg atg ttt ttg tct cta gga gtt ggg gcg gat caa gga tgc gcc atc 2470
Met Met Phe Leu Ser Leu Gly Val Gly Ala Asp Gln Gly Cys Ala Ile
770 775 780

aac ttt ggc aag aga gag ctc aag tgc gga gat ggt atc ttc ata ttt 2518
Asn Phe Gly Lys Arg Glu Leu Lys Cys Gly Asp Gly Ile Phe Ile Phe
785 790 795 800

aga gac tct gat gac tgg ctg aac aag tac tca tac tat cca gaa gat 2566
Arg Asp Ser Asp Asp Trp Leu Asn Lys Tyr Ser Tyr Tyr Pro Glu Asp
805 810 815

cct gtg aag ctt gca tca ata gtg aaa gcc tct ttt gaa gaa ggg aag 2614
Pro Val Lys Leu Ala Ser Ile Val Lys Ala Ser Phe Glu Glu Gly Lys
820 825 830

tgt ggc cta aat tca gtt gac tcc ctt gag cat gag atg tgg aga agc 2662
Cys Gly Leu Asn Ser Val Asp Ser Leu Glu His Glu Met Trp Arg Ser
835 840 845

agg gca gat gag atc aat gcc att ctt gag gaa aac gag gtg gac att 2710
Arg Ala Asp Glu Ile Asn Ala Ile Leu Glu Glu Asn Glu Val Asp Ile
850 855 860

tct gtt gtc gtg cag gat cca aag aat gtt tac cag aga gga act cat 2758
Ser Val Val Val Gln Asp Pro Lys Asn Val Tyr Gln Arg Gly Thr His
865 870 875 880

cca ttt tcc aga att cgg gac ggt ctg cag tat ggt tgg aag act tgg 2806
Pro Phe Ser Arg Ile Arg Asp Gly Leu Gln Tyr Gly Trp Lys Thr Trp
885 890 895

ggt aag aac ctt gtg ttc tcc cca ggg agg aag aat gga agc ttc atc 2854
Gly Lys Asn Leu Val Phe Ser Pro Gly Arg Lys Asn Gly Ser Phe Ile
900 905 910

ata gat gga aag tcc agg aaa gaa tgc ccg ttt tca aac cgg gtc tgg 2902

Ile Asp Gly Lys Ser Arg Lys Glu Cys Pro Phe Ser Asn Arg Val Trp
915 920 925

aat tct ttc cag ata gag gag ttt ggg acg gga gtg ttc acc aca cgc 2950
Asn Ser Phe Gln Ile Glu Glu Phe Gly Thr Gly Val Phe Thr Thr Arg
930 935 940

gtg tac atg gac gca gtc ttt gaa tac acc ata gac tgc gat gga tct 2998
Val Tyr Met Asp Ala Val Phe Glu Tyr Thr Ile Asp Cys Asp Gly Ser
945 950 955 960

atc ttg ggt gca gcg gtg aac gga aaa aag agt gcc cat ggc tct cca 3046
Ile Leu Gly Ala Ala Val Asn Gly Lys Lys Ser Ala His Gly Ser Pro
965 970 975

aca ttt tgg atg gga agt cat gaa gta aat ggg aca tgg atg atc cac 3094
Thr Phe Trp Met Gly Ser His Glu Val Asn Gly Thr Trp Met Ile His
980 985 990

acc ttg gag gca tta gat tac aag gag tgt gag tgg cca ctg aca cat 3142
Thr Leu Glu Ala Leu Asp Tyr Lys Glu Cys Glu Trp Pro Leu Thr His
995 1000 1005

acg att gga aca tca gtt gaa gag agt gaa atg ttc atg ccg aga tca 3190
Thr Ile Gly Thr Ser Val Glu Glu Ser Glu Met Phe Met Pro Arg Ser
1010 1015 1020

atc gga ggc cca gtt agc tct cac aat cat atc cct gga tac aag gtt 3238
Ile Gly Gly Pro Val Ser Ser His Asn His Ile Pro Gly Tyr Lys Val
1025 1030 1035 1040

cag acg aac gga cct tgg atg cag gta cca cta gaa gtg aag aga gaa 3286
Gln Thr Asn Gly Pro Trp Met Gln Val Pro Leu Glu Val Lys Arg Glu
1045 1050 1055

gct tgc cca ggg act agc gtg atc att gat ggc aac tgt gat gga cgg 3334
Ala Cys Pro Gly Thr Ser Val Ile Ile Asp Gly Asn Cys Asp Gly Arg
1060 1065 1070

gga aaa tca acc aga tcc acc acg gat agc ggg aaa att att cct gaa 3382
Gly Lys Ser Thr Arg Ser Thr Thr Asp Ser Gly Lys Ile Ile Pro Glu
1075 1080 1085

tgg tgt tgc cgc tcc tgc aca atg ccg cct gtg agc ttc cat ggt agt 3430
Trp Cys Cys Arg Ser Cys Thr Met Pro Pro Val Ser Phe His Gly Ser
1090 1095 1100

gat ggg tgt tgg tat ccc atg gaa att agg cca agg aaa acg cat gaa 3478
Asp Gly Cys Trp Tyr Pro Met Glu Ile Arg Pro Arg Lys Thr His Glu
1105 1110 1115 1120

agc cat ctg gtg cgc tcc tgg gtt aca gct gga gaa ata cat gct gtc 3526

```
Ser His Leu Val Arg Ser Trp Val Thr Ala Gly Glu Ile His Ala Val
                1125                1130                1135

cct ttt ggt ttg gtg agc atg atg ata gca atg gaa gtg gtc cta agg    3574
Pro Phe Gly Leu Val Ser Met Met Ile Ala Met Glu Val Val Leu Arg
            1140                1145                1150

aaa aga cag gga cca aag caa atg ttg gtt gga gga gtg gtg ctc ttg    3622
Lys Arg Gln Gly Pro Lys Gln Met Leu Val Gly Gly Val Val Leu Leu
        1155                1160                1165

gga gca atg ctg gtc ggg caa gta act ctc ctt gat ttg ctg aaa ctc    3670
Gly Ala Met Leu Val Gly Gln Val Thr Leu Leu Asp Leu Leu Lys Leu
    1170                1175                1180

aca gtg gct gtg gga ttg cat ttc cat gag atg aac aat gga gga gac    3718
Thr Val Ala Val Gly Leu His Phe His Glu Met Asn Asn Gly Gly Asp
1185                1190                1195                1200

gcc atg tat atg gcg ttg att gct gcc ttt tca atc aga cca ggg ctg    3766
Ala Met Tyr Met Ala Leu Ile Ala Ala Phe Ser Ile Arg Pro Gly Leu
                1205                1210                1215

ctc atc ggc ttt ggg ctc agg acc cta tgg agc cct cgg gaa cgc ctt    3814
Leu Ile Gly Phe Gly Leu Arg Thr Leu Trp Ser Pro Arg Glu Arg Leu
            1220                1225                1230

gta ctg gcc cta gga gca gcc atg gtg gag att gcc ttg ggt ggc atg    3862
Val Leu Ala Leu Gly Ala Ala Met Val Glu Ile Ala Leu Gly Gly Met
        1235                1240                1245

atg ggc ggc ctg tgg aag tat cta aat gca gtt tct ctc tgc atc ctg    3910
Met Gly Gly Leu Trp Lys Tyr Leu Asn Ala Val Ser Leu Cys Ile Leu
    1250                1255                1260

aca ata aat gct gta gct tct agg aaa gca tca aat acc atc ttg ccc    3958
Thr Ile Asn Ala Val Ala Ser Arg Lys Ala Ser Asn Thr Ile Leu Pro
1265                1270                1275                1280

ctc atg gct ctg ttg aca cct gtc act atg gct gag gtg aga ctt gcc    4006
Leu Met Ala Leu Leu Thr Pro Val Thr Met Ala Glu Val Arg Leu Ala
                1285                1290                1295

aca atg ctc ttt tgt acc gtg gtt atc ata ggg gtc ctt cac cag aac    4054
Thr Met Leu Phe Cys Thr Val Val Ile Ile Gly Val Leu His Gln Asn
            1300                1305                1310

tcc aag gac acc tcc atg cag aag act ata cct ctg gtg gcc ctc aca    4102
Ser Lys Asp Thr Ser Met Gln Lys Thr Ile Pro Leu Val Ala Leu Thr
        1315                1320                1325

ctc aca tct tac ctg ggc ttg aca caa cct ttt ttg ggc ctg tgt gca    4150
```

Leu Thr Ser Tyr Leu Gly Leu Thr Gln Pro Phe Leu Gly Leu Cys Ala
1330 1335 1340

ttt ctg gca acc cgc ata ttt ggg cga agg agt atc cca gtg aat gag 4198
Phe Leu Ala Thr Arg Ile Phe Gly Arg Arg Ser Ile Pro Val Asn Glu
1345 1350 1355 1360

gca ctc gca gca gct ggt cta gtg gga gtg ctg gca gga ctg gct ttt 4246
Ala Leu Ala Ala Ala Gly Leu Val Gly Val Leu Ala Gly Leu Ala Phe
1365 1370 1375

cag gag atg gag aac ttc ctt ggt ccg att gca gtt gga gga atc ctg 4294
Gln Glu Met Glu Asn Phe Leu Gly Pro Ile Ala Val Gly Gly Ile Leu
1380 1385 1390

atg atg ctg gtt agc gtg gct ggg agg gtg gat ggg cta gag ctc aag 4342
Met Met Leu Val Ser Val Ala Gly Arg Val Asp Gly Leu Glu Leu Lys
1395 1400 1405

aag ctt ggt gaa gtt tca tgg gaa gag gag gcg gag atc agc gga agt 4390
Lys Leu Gly Glu Val Ser Trp Glu Glu Glu Ala Glu Ile Ser Gly Ser
1410 1415 1420

tcc gcc cgc tat gat gtg gca ctc agt gaa caa ggg gag ttc aag ctg 4438
Ser Ala Arg Tyr Asp Val Ala Leu Ser Glu Gln Gly Glu Phe Lys Leu
1425 1430 1435 1440

ctt tct gaa gag aaa gtg cca tgg gac cag gtt gtg atg acc tcg ctg 4486
Leu Ser Glu Glu Lys Val Pro Trp Asp Gln Val Val Met Thr Ser Leu
1445 1450 1455

gcc ttg gtt ggg gct gcc att cat cca ttt gct ctt ctg ctg gtc ctt 4534
Ala Leu Val Gly Ala Ala Ile His Pro Phe Ala Leu Leu Leu Val Leu
1460 1465 1470

gct ggg tgg ctg ttt cat gtc agg gga gct agg aga agt ggg gat gtc 4582
Ala Gly Trp Leu Phe His Val Arg Gly Ala Arg Arg Ser Gly Asp Val
1475 1480 1485

ttg tgg gat att ccc act cct aag atc att gag gaa tgt gaa cat ctg 4630
Leu Trp Asp Ile Pro Thr Pro Lys Ile Ile Glu Glu Cys Glu His Leu
1490 1495 1500

gag gat ggg att tat ggc ata ttc cag tca acc ttc ttg ggg gcc tcc 4678
Glu Asp Gly Ile Tyr Gly Ile Phe Gln Ser Thr Phe Leu Gly Ala Ser
1505 1510 1515 1520

cag cga gga gtg gga gtg gca cag gga ggg gtg ttc cac aca atg tgg 4726
Gln Arg Gly Val Gly Val Ala Gln Gly Gly Val Phe His Thr Met Trp
1525 1530 1535

cat gtc aca aga gga gct ttc ctt gtc agg aat ggc aag aag ttg att 4774

His Val Thr Arg Gly Ala Phe Leu Val Arg Asn Gly Lys Lys Leu Ile
1540 1545 1550

cca tct tgg gct tca gta aag gaa gac ctt gtc gcc tat ggt ggc tca 4822
Pro Ser Trp Ala Ser Val Lys Glu Asp Leu Val Ala Tyr Gly Gly Ser
1555 1560 1565

tgg aag ttg gaa ggc aga tgg gat gga gag gaa gag gtc caa ttg atc 4870
Trp Lys Leu Glu Gly Arg Trp Asp Gly Glu Glu Glu Val Gln Leu Ile
1570 1575 1580

gct gct gtt cca gga aag aac gtg gtc aac gtc cag aca aaa ccg agc 4918
Ala Ala Val Pro Gly Lys Asn Val Val Asn Val Gln Thr Lys Pro Ser
1585 1590 1595 1600

ttg ttc aaa gtg agg aat ggg gga gaa atc ggg gct gtc gct ctt gac 4966
Leu Phe Lys Val Arg Asn Gly Gly Glu Ile Gly Ala Val Ala Leu Asp
1605 1610 1615

tat ccg agt ggc act tca gga tct cct att gtt aac agg aac gga gag 5014
Tyr Pro Ser Gly Thr Ser Gly Ser Pro Ile Val Asn Arg Asn Gly Glu
1620 1625 1630

gtg att ggg ctg tac ggc aat ggc atc ctt gtc ggt gac aac tcc ttc 5062
Val Ile Gly Leu Tyr Gly Asn Gly Ile Leu Val Gly Asp Asn Ser Phe
1635 1640 1645

gtg tcc gcc ata tcc cag act gag gtg aag gaa gaa gga aag gag gag 5110
Val Ser Ala Ile Ser Gln Thr Glu Val Lys Glu Glu Gly Lys Glu Glu
1650 1655 1660

ctc caa gag atc ccg aca atg cta aag aaa gga atg aca act atc ctt 5158
Leu Gln Glu Ile Pro Thr Met Leu Lys Lys Gly Met Thr Thr Ile Leu
1665 1670 1675 1680

gat ttt cat cct gga gct ggg aag aca aga cgt ttt ctc cca cag atc 5206
Asp Phe His Pro Gly Ala Gly Lys Thr Arg Arg Phe Leu Pro Gln Ile
1685 1690 1695

ttg gcc gag tgc gca cgg aga cgc ttg cgc act ctt gtg ttg gcc ccc 5254
Leu Ala Glu Cys Ala Arg Arg Arg Leu Arg Thr Leu Val Leu Ala Pro
1700 1705 1710

acc agg gtt gtt ctt tct gaa atg aag gag gct ttt cac ggc ctg gac 5302
Thr Arg Val Val Leu Ser Glu Met Lys Glu Ala Phe His Gly Leu Asp
1715 1720 1725

gtg aaa ttc cac aca cag gct ttt tcc gct cac ggc agc ggg aga gaa 5350
Val Lys Phe His Thr Gln Ala Phe Ser Ala His Gly Ser Gly Arg Glu
1730 1735 1740

gtc att gat gcc atg tgc cat gcc acc cta act tac agg atg ttg gaa 5398

Val Ile Asp Ala Met Cys His Ala Thr Leu Thr Tyr Arg Met Leu Glu
1745 1750 1755 1760

cca act agg gtt gtt aac tgg gaa gtg atc atc atg gat gaa gcc cat 5446
Pro Thr Arg Val Val Asn Trp Glu Val Ile Ile Met Asp Glu Ala His
1765 1770 1775

ttt ttg gat cca gct agc ata gcc gcc aga ggt tgg gca gcg cac aga 5494
Phe Leu Asp Pro Ala Ser Ile Ala Ala Arg Gly Trp Ala Ala His Arg
1780 1785 1790

gct agg gca aat gaa agt gca aca atc ttg atg aca gcc aca ccg cct 5542
Ala Arg Ala Asn Glu Ser Ala Thr Ile Leu Met Thr Ala Thr Pro Pro
1795 1800 1805

ggg act agt gat gaa ttt cca cat tca aat ggt gaa ata gaa gat gtt 5590
Gly Thr Ser Asp Glu Phe Pro His Ser Asn Gly Glu Ile Glu Asp Val
1810 1815 1820

caa acg gac ata ccc agt gag ccc tgg aac aca ggg cat gac tgg atc 5638
Gln Thr Asp Ile Pro Ser Glu Pro Trp Asn Thr Gly His Asp Trp Ile
1825 1830 1835 1840

ctg gct gac aaa agg ccc acg gca tgg ttc ctt cca tcc atc aga gct 5686
Leu Ala Asp Lys Arg Pro Thr Ala Trp Phe Leu Pro Ser Ile Arg Ala
1845 1850 1855

gca aat gtc atg gct gcc tct ttg cgt aag gct gga aag agt gtg gtg 5734
Ala Asn Val Met Ala Ala Ser Leu Arg Lys Ala Gly Lys Ser Val Val
1860 1865 1870

gtc ctg aac agg aaa acc ttt gag aga gaa tac ccc acg ata aag cag 5782
Val Leu Asn Arg Lys Thr Phe Glu Arg Glu Tyr Pro Thr Ile Lys Gln
1875 1880 1885

aag aaa cct gac ttt ata ttg gcc act gac ata gct gaa atg gga gcc 5830
Lys Lys Pro Asp Phe Ile Leu Ala Thr Asp Ile Ala Glu Met Gly Ala
1890 1895 1900

aac ctt tgc gtg gag cga gtg ctg gat tgc agg acg gct ttt aag cct 5878
Asn Leu Cys Val Glu Arg Val Leu Asp Cys Arg Thr Ala Phe Lys Pro
1905 1910 1915 1920

gtg ctt gtg gat gaa ggg agg aag gtg gca ata aaa ggg cca ctt cgc 5926
Val Leu Val Asp Glu Gly Arg Lys Val Ala Ile Lys Gly Pro Leu Arg
1925 1930 1935

atc tcc gca tcc tct gct gct caa agg agg ggg cgc att ggg aga aat 5974
Ile Ser Ala Ser Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn
1940 1945 1950

ccc aac aga gat gga gac tca tac tac tat tct gag cct aca agt gaa 6022

Pro Asn Arg Asp Gly Asp Ser Tyr Tyr Tyr Ser Glu Pro Thr Ser Glu
1955 1960 1965

gat aat gcc cac cac gtc tgc tgg ttg gag gcc tca atg ctc ttg gac 6070
Asp Asn Ala His His Val Cys Trp Leu Glu Ala Ser Met Leu Leu Asp
1970 1975 1980

aac atg gag gtg agg ggt gga atg gtc gcc cca ctc tat ggc gtt gaa 6118
Asn Met Glu Val Arg Gly Gly Met Val Ala Pro Leu Tyr Gly Val Glu
1985 1990 1995 2000

gga act aaa aca cca gtt tcc cct ggt gaa atg aga ctg agg gat gac 6166
Gly Thr Lys Thr Pro Val Ser Pro Gly Glu Met Arg Leu Arg Asp Asp
2005 2010 2015

cag agg aaa gtc ttc aga gaa cta gtg agg aat tgt gac ctg ccc gtt 6214
Gln Arg Lys Val Phe Arg Glu Leu Val Arg Asn Cys Asp Leu Pro Val
2020 2025 2030

tgg ctt tcg tgg caa gtg gcc aag gct ggt ttg aag acg aat gat cgt 6262
Trp Leu Ser Trp Gln Val Ala Lys Ala Gly Leu Lys Thr Asn Asp Arg
2035 2040 2045

aag tgg tgt ttt gaa ggc cct gag gaa cat gag atc ttg aat gac agc 6310
Lys Trp Cys Phe Glu Gly Pro Glu Glu His Glu Ile Leu Asn Asp Ser
2050 2055 2060

ggt gaa aca gtg aag tgc agg gct cct gga gga gca aag aag cct ctg 6358
Gly Glu Thr Val Lys Cys Arg Ala Pro Gly Gly Ala Lys Lys Pro Leu
2065 2070 2075 2080

cgc cca agg tgg tgt gat gaa agg gtg tca tct gac cag agt gcg ctg 6406
Arg Pro Arg Trp Cys Asp Glu Arg Val Ser Ser Asp Gln Ser Ala Leu
2085 2090 2095

tct gaa ttt att aag ttt gct gaa ggt agg agg gga gct gcg gaa gtg 6454
Ser Glu Phe Ile Lys Phe Ala Glu Gly Arg Arg Gly Ala Ala Glu Val
2100 2105 2110

cta gtt gtg ctg agt gaa ctc cct gat ttc ctg gct aaa aaa ggt gga 6502
Leu Val Val Leu Ser Glu Leu Pro Asp Phe Leu Ala Lys Lys Gly Gly
2115 2120 2125

gag gca atg gat acc atc agt gtg ttt ctc cac tct gag gaa ggc tct 6550
Glu Ala Met Asp Thr Ile Ser Val Phe Leu His Ser Glu Glu Gly Ser
2130 2135 2140

agg gct tac cgc aat gca cta tca atg atg cct gag gca atg aca ata 6598
Arg Ala Tyr Arg Asn Ala Leu Ser Met Met Pro Glu Ala Met Thr Ile
2145 2150 2155 2160

gtc atg ctg ttt ata ctg gct gga cta ctg aca tcg gga atg gtc atc 6646

Val Met Leu Phe Ile Leu Ala Gly Leu Leu Thr Ser Gly Met Val Ile
2165 2170 2175

ttt ttc atg tct ccc aaa ggc atc agt aga atg tct atg gcg atg ggc 6694
Phe Phe Met Ser Pro Lys Gly Ile Ser Arg Met Ser Met Ala Met Gly
2180 2185 2190

aca atg gcc ggc tgt gga tat ctc atg ttc ctt gga ggc gtc aaa ccc 6742
Thr Met Ala Gly Cys Gly Tyr Leu Met Phe Leu Gly Gly Val Lys Pro
2195 2200 2205

act cac atc tcc tat atc atg ctc ata ttc ttt gtc ctg atg gtg gtt 6790
Thr His Ile Ser Tyr Ile Met Leu Ile Phe Phe Val Leu Met Val Val
2210 2215 2220

gtg atc ccc gag cca ggg caa caa agg tcc atc caa gac aac caa gtg 6838
Val Ile Pro Glu Pro Gly Gln Gln Arg Ser Ile Gln Asp Asn Gln Val
2225 2230 2235 2240

gca tac ctc att att ggc atc ctg acg ctg gtt tca gtg gtg gca gcc 6886
Ala Tyr Leu Ile Ile Gly Ile Leu Thr Leu Val Ser Val Val Ala Ala
2245 2250 2255

aac gag cta ggc atg ctg gag aaa acc aaa gag gac ctc ttt ggg aag 6934
Asn Glu Leu Gly Met Leu Glu Lys Thr Lys Glu Asp Leu Phe Gly Lys
2260 2265 2270

aag aac tta att cca tct agt gct tca ccc tgg agt tgg ccg gat ctt 6982
Lys Asn Leu Ile Pro Ser Ser Ala Ser Pro Trp Ser Trp Pro Asp Leu
2275 2280 2285

gac ctg aag cca gga gct gcc tgg aca gtg tac gtt ggc att gtt aca 7030
Asp Leu Lys Pro Gly Ala Ala Trp Thr Val Tyr Val Gly Ile Val Thr
2290 2295 2300

atg ctc tct cca atg ttg cac cac tgg atc aaa gtc gaa tat ggc aac 7078
Met Leu Ser Pro Met Leu His His Trp Ile Lys Val Glu Tyr Gly Asn
2305 2310 2315 2320

ctg tct ctg tct gga ata gcc cag tca gcc tca gtc ctt tct ttc atg 7126
Leu Ser Leu Ser Gly Ile Ala Gln Ser Ala Ser Val Leu Ser Phe Met
2325 2330 2335

gac aag ggg ata cca ttc atg aag atg aat atc tcg gtc ata ata ctg 7174
Asp Lys Gly Ile Pro Phe Met Lys Met Asn Ile Ser Val Ile Ile Leu
2340 2345 2350

ctg atc agt ggc tgg aat tca ata aca gtg atg cct ctg ctc tgt ggc 7222
Leu Ile Ser Gly Trp Asn Ser Ile Thr Val Met Pro Leu Leu Cys Gly
2355 2360 2365

ata ggg tgc gcc atg ctc cac tgg tct ctc att tta cct gga atc aaa 7270

Ile Gly Cys Ala Met Leu His Trp Ser Leu Ile Leu Pro Gly Ile Lys
2370 2375 2380

gcg cag cag tca aag ctt gca cag aga agg gtg ttc cat ggc gtt gcc 7318
Ala Gln Gln Ser Lys Leu Ala Gln Arg Arg Val Phe His Gly Val Ala
2385 2390 2395 2400

aag aac cct gtg gtt gat ggg aat cca aca gtt gac att gag gaa gct 7366
Lys Asn Pro Val Val Asp Gly Asn Pro Thr Val Asp Ile Glu Glu Ala
2405 2410 2415

cct gaa atg cct gcc ctt tat gag aag aaa ctg gct cta tat ctc ctt 7414
Pro Glu Met Pro Ala Leu Tyr Glu Lys Lys Leu Ala Leu Tyr Leu Leu
2420 2425 2430

ctt gct ctc agc cta gct tct gtt gcc atg tgc aga acg ccc ttt tca 7462
Leu Ala Leu Ser Leu Ala Ser Val Ala Met Cys Arg Thr Pro Phe Ser
2435 2440 2445

ttg gct gaa ggc att gtc cta gca tca gct gcc tta ggg ccg ctc ata 7510
Leu Ala Glu Gly Ile Val Leu Ala Ser Ala Ala Leu Gly Pro Leu Ile
2450 2455 2460

gag gga aac acc agc ctt ctt tgg aat gga ccc atg gct gtc tcc atg 7558
Glu Gly Asn Thr Ser Leu Leu Trp Asn Gly Pro Met Ala Val Ser Met
2465 2470 2475 2480

aca gga gtc atg cgg ggg aat tac tat gct ttt gtg gga gtc atg tac 7606
Thr Gly Val Met Arg Gly Asn Tyr Tyr Ala Phe Val Gly Val Met Tyr
2485 2490 2495

aat cta tgg aag atg aaa act gga cgc cgg ggg agt gcg aat gga aaa 7654
Asn Leu Trp Lys Met Lys Thr Gly Arg Arg Gly Ser Ala Asn Gly Lys
2500 2505 2510

act ttg ggt gaa gtc tgg aag agg gaa ctg aat ctg ttg gac aag caa 7702
Thr Leu Gly Glu Val Trp Lys Arg Glu Leu Asn Leu Leu Asp Lys Gln
2515 2520 2525

cag ttt gag ttg tat aaa agg acc gac att gtg gag gtg gat cgt gat 7750
Gln Phe Glu Leu Tyr Lys Arg Thr Asp Ile Val Glu Val Asp Arg Asp
2530 2535 2540

acg gca cgc agg cat ttg gcc gaa ggg aag gtg gac acc ggg gtg gcg 7798
Thr Ala Arg Arg His Leu Ala Glu Gly Lys Val Asp Thr Gly Val Ala
2545 2550 2555 2560

gtc tcc agg ggg acc gca aag tta agg tgg ttc cat gag cgt ggc tat 7846
Val Ser Arg Gly Thr Ala Lys Leu Arg Trp Phe His Glu Arg Gly Tyr
2565 2570 2575

gtc aag ctg gaa ggt agg gtg att gac ctg ggg tgt ggc cgc gga ggc 7894

Val Lys Leu Glu Gly Arg Val Ile Asp Leu Gly Cys Gly Arg Gly Gly
2580 2585 2590

tgg tgt tac tac gct gct gcg caa aag gaa gtg agt ggg gtc aaa gga 7942
Trp Cys Tyr Tyr Ala Ala Ala Gln Lys Glu Val Ser Gly Val Lys Gly
2595 2600 2605

ttc act ctt gga aga gac ggc cat gag aaa ccc atg aat gtg caa agt 7990
Phe Thr Leu Gly Arg Asp Gly His Glu Lys Pro Met Asn Val Gln Ser
2610 2615 2620

ctg gga tgg aac atc att acc ttc aag gac aaa act gat atc cac cgc 8038
Leu Gly Trp Asn Ile Ile Thr Phe Lys Asp Lys Thr Asp Ile His Arg
2625 2630 2635 2640

cta gaa cca gtg aaa tgt gac acc ctt ttg tgt gac att gga gag tca 8086
Leu Glu Pro Val Lys Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser
2645 2650 2655

tca tcg tca tcg gtc aca gag ggg gaa agg acc gtg aga gtt ctt gat 8134
Ser Ser Ser Ser Val Thr Glu Gly Glu Arg Thr Val Arg Val Leu Asp
2660 2665 2670

act gta gaa aaa tgg ctg gct tgt ggg gtt gac aac ttc tgt gtg aag 8182
Thr Val Glu Lys Trp Leu Ala Cys Gly Val Asp Asn Phe Cys Val Lys
2675 2680 2685

gtg tta gct cca tac atg cca gat gtt ctc gag aaa ctg gaa ttg ctc 8230
Val Leu Ala Pro Tyr Met Pro Asp Val Leu Glu Lys Leu Glu Leu Leu
2690 2695 2700

caa agg agg ttt ggc gga aca gtg atc agg aac cct ctc tcc agg aat 8278
Gln Arg Arg Phe Gly Gly Thr Val Ile Arg Asn Pro Leu Ser Arg Asn
2705 2710 2715 2720

tcc act cat gaa atg tac tac gtg tct gga gcc cgc agc aat gtc aca 8326
Ser Thr His Glu Met Tyr Tyr Val Ser Gly Ala Arg Ser Asn Val Thr
2725 2730 2735

ttt act gtg aac caa aca tcc cgc ctc ctg atg agg aga atg agg cgt 8374
Phe Thr Val Asn Gln Thr Ser Arg Leu Leu Met Arg Arg Met Arg Arg
2740 2745 2750

cca act gga aaa gtg acc ctg gag gct gac gtc atc ctc cca att ggg 8422
Pro Thr Gly Lys Val Thr Leu Glu Ala Asp Val Ile Leu Pro Ile Gly
2755 2760 2765

aca cgc agt gtt gag aca gac aag gga ccc ctg gac aaa gag gcc ata 8470
Thr Arg Ser Val Glu Thr Asp Lys Gly Pro Leu Asp Lys Glu Ala Ile
2770 2775 2780

gaa gaa agg gtt gag agg ata aaa tct gag tac atg acc tct tgg ttt 8518

Glu Glu Arg Val Glu Arg Ile Lys Ser Glu Tyr Met Thr Ser Trp Phe
2785 2790 2795 2800

tat gac aat gac aac ccc tac agg acc tgg cac tac tgt ggc tcc tat 8566
Tyr Asp Asn Asp Asn Pro Tyr Arg Thr Trp His Tyr Cys Gly Ser Tyr
2805 2810 2815

gtc aca aaa acc tca gga agt gcg gcg agc atg gta aat ggt gtt att 8614
Val Thr Lys Thr Ser Gly Ser Ala Ala Ser Met Val Asn Gly Val Ile
2820 2825 2830

aaa att ctg aca tac cca tgg gac agg ata gag gag gtc aca aga atg 8662
Lys Ile Leu Thr Tyr Pro Trp Asp Arg Ile Glu Glu Val Thr Arg Met
2835 2840 2845

gca atg act gac aca acc cct ttt gga cag caa aga gtg ttt aaa gaa 8710
Ala Met Thr Asp Thr Thr Pro Phe Gly Gln Gln Arg Val Phe Lys Glu
2850 2855 2860

aaa gtt gac acc aga gca aag gat cca cca gcg gga act agg aag atc 8758
Lys Val Asp Thr Arg Ala Lys Asp Pro Pro Ala Gly Thr Arg Lys Ile
2865 2870 2875 2880

atg aaa gtt gtc aac agg tgg ctg ttc cgc cac ctg gcc aga gaa aag 8806
Met Lys Val Val Asn Arg Trp Leu Phe Arg His Leu Ala Arg Glu Lys
2885 2890 2895

aac ccc aga ctg tgc aca aag gaa gaa ttt att gca aaa gtc cga agt 8854
Asn Pro Arg Leu Cys Thr Lys Glu Glu Phe Ile Ala Lys Val Arg Ser
2900 2905 2910

cat gca gcc att gga gct tac ctg gaa gaa caa gaa cag tgg aag act 8902
His Ala Ala Ile Gly Ala Tyr Leu Glu Glu Gln Glu Gln Trp Lys Thr
2915 2920 2925

gcc aat gag gct gtt caa gac cca aag ttc tgg gaa ctg gtg gat gaa 8950
Ala Asn Glu Ala Val Gln Asp Pro Lys Phe Trp Glu Leu Val Asp Glu
2930 2935 2940

gaa agg aag ctg cac caa caa ggc agg tgt cgg act tgt gtg tac aac 8998
Glu Arg Lys Leu His Gln Gln Gly Arg Cys Arg Thr Cys Val Tyr Asn
2945 2950 2955 2960

atg atg ggg aaa aga gag aag aag ctg tca gag ttt ggg aaa gca aag 9046
Met Met Gly Lys Arg Glu Lys Lys Leu Ser Glu Phe Gly Lys Ala Lys
2965 2970 2975

gga agc cgt gcc ata tgg tat atg tgg ctg gga gcg cgg tat ctt gag 9094
Gly Ser Arg Ala Ile Trp Tyr Met Trp Leu Gly Ala Arg Tyr Leu Glu
2980 2985 2990

ttt gag gcc ctg gga ttc ctg aat gag gac cat tgg gct tcc agg gaa 9142

Phe Glu Ala Leu Gly Phe Leu Asn Glu Asp His Trp Ala Ser Arg Glu
2995 3000 3005

aac tca gga gga gga gtg gaa ggc att ggc tta caa tac cta gga tat 9190
Asn Ser Gly Gly Gly Val Glu Gly Ile Gly Leu Gln Tyr Leu Gly Tyr
3010 3015 3020

gtg atc aga gac ctg gct gca atg gat ggt ggt gga ttc tac gcg gat 9238
Val Ile Arg Asp Leu Ala Ala Met Asp Gly Gly Gly Phe Tyr Ala Asp
3025 3030 3035 3040

gac acc gct gga tgg gac acg cgc atc aca gag gca gac ctt gat gat 9286
Asp Thr Ala Gly Trp Asp Thr Arg Ile Thr Glu Ala Asp Leu Asp Asp
3045 3050 3055

gaa cag gag atc ttg aac tac atg agc cca cat cac aaa aaa ctg gca 9334
Glu Gln Glu Ile Leu Asn Tyr Met Ser Pro His His Lys Lys Leu Ala
3060 3065 3070

caa gca gtg atg gaa atg aca tac aag aac aaa gtg gtg aaa gtg ttg 9382
Gln Ala Val Met Glu Met Thr Tyr Lys Asn Lys Val Val Lys Val Leu
3075 3080 3085

aga cca gcc cca gga ggg aaa gcc tac atg gat gtc ata agt cga cga 9430
Arg Pro Ala Pro Gly Gly Lys Ala Tyr Met Asp Val Ile Ser Arg Arg
3090 3095 3100

gac cag aga gga tcc ggg cag gta gtg act tat gct ctg aac acc atc 9478
Asp Gln Arg Gly Ser Gly Gln Val Val Thr Tyr Ala Leu Asn Thr Ile
3105 3110 3115 3120

acc aac ttg aaa gtc caa ttg atc aga atg gca gaa gca gag atg gtg 9526
Thr Asn Leu Lys Val Gln Leu Ile Arg Met Ala Glu Ala Glu Met Val
3125 3130 3135

ata cat cac caa cat gtt caa gat tgt gat gaa tca gtt ctg acc agg 9574
Ile His His Gln His Val Gln Asp Cys Asp Glu Ser Val Leu Thr Arg
3140 3145 3150

ctg gag gca tgg ctc act gag cac gga tgt aac aga ctg aag agg atg 9622
Leu Glu Ala Trp Leu Thr Glu His Gly Cys Asn Arg Leu Lys Arg Met
3155 3160 3165

gcg gtg agt gga gac gac tgt gtg gtc cgg ccc atc gat gac agg ttc 9670
Ala Val Ser Gly Asp Asp Cys Val Val Arg Pro Ile Asp Asp Arg Phe
3170 3175 3180

ggc ctg gcc ctg tcc cat ctc aac gcc atg tcc aag gtt aga aag gac 9718
Gly Leu Ala Leu Ser His Leu Asn Ala Met Ser Lys Val Arg Lys Asp
3185 3190 3195 3200

ata tct gaa tgg cag cca tca aaa ggg tgg aat gat tgg gag aat gtg 9766

Ile Ser Glu Trp Gln Pro Ser Lys Gly Trp Asn Asp Trp Glu Asn Val
3205 3210 3215

ccc ttc tgt tcc cac cac ttc cat gaa cta cag ctg aag gat ggc agg 9814
Pro Phe Cys Ser His His Phe His Glu Leu Gln Leu Lys Asp Gly Arg
3220 3225 3230

agg att gtg gtg cct tgc cga gaa cag gac gag ctc att ggg aga gga 9862
Arg Ile Val Val Pro Cys Arg Glu Gln Asp Glu Leu Ile Gly Arg Gly
3235 3240 3245

agg gtg tct cca gga aac ggc tgg atg atc aag gaa aca gct tgc ctc 9910
Arg Val Ser Pro Gly Asn Gly Trp Met Ile Lys Glu Thr Ala Cys Leu
3250 3255 3260

agc aaa gcc tat gcc aac atg tgg tca ctg atg tat ttt cac aaa agg 9958
Ser Lys Ala Tyr Ala Asn Met Trp Ser Leu Met Tyr Phe His Lys Arg
3265 3270 3275 3280

gac atg agg cta ctg tca ttg gct gtt tcc tca gct gtt ccc acc tca 10006
Asp Met Arg Leu Leu Ser Leu Ala Val Ser Ser Ala Val Pro Thr Ser
3285 3290 3295

tgg gtt cca caa gga cgc aca aca tgg tcg att cat ggg aaa ggg gag 10054
Trp Val Pro Gln Gly Arg Thr Thr Trp Ser Ile His Gly Lys Gly Glu
3300 3305 3310

tgg atg acc acg gaa gac atg ctt gag gtg tgg aac aga gta tgg ata 10102
Trp Met Thr Thr Glu Asp Met Leu Glu Val Trp Asn Arg Val Trp Ile
3315 3320 3325

acc aac aac cca cac atg cag gac aag aca atg gtg aaa gaa tgg aga 10150
Thr Asn Asn Pro His Met Gln Asp Lys Thr Met Val Lys Glu Trp Arg
3330 3335 3340

gat gtc cct tat cta acc aag aga caa gac aag ctg tgc gga tca ctg 10198
Asp Val Pro Tyr Leu Thr Lys Arg Gln Asp Lys Leu Cys Gly Ser Leu
3345 3350 3355 3360

att gga atg acc aat agg gcc acc tgg gcc tcc cac atc cat ttg gtc 10246
Ile Gly Met Thr Asn Arg Ala Thr Trp Ala Ser His Ile His Leu Val
3365 3370 3375

atc cat cgt atc cga acg ctg att gga cag gag aaa tat act gac tac 10294
Ile His Arg Ile Arg Thr Leu Ile Gly Gln Glu Lys Tyr Thr Asp Tyr
3380 3385 3390

cta aca gtc atg gac aga tat tct gtg gat gct gac ctg caa ccg ggt 10342
Leu Thr Val Met Asp Arg Tyr Ser Val Asp Ala Asp Leu Gln Pro Gly
3395 3400 3405

gag ctt atc tga aac acc atc taa tag gaa taa ccg gga tac aaa cca 10390

Glu Leu Ile Asn Thr Ile Glu Pro Gly Tyr Lys Pro
3410 3415 3420

cgg gtg gag aac cgg act ccc cac aac ttg aaa ccg gga tat aaa cca 10438
Arg Val Glu Asn Arg Thr Pro His Asn Leu Lys Pro Gly Tyr Lys Pro
3425 3430 3435 3440

cgg ctg gag aac cgg act ccg cac tta aaa tga aac aga aac cgg gat 10486
Arg Leu Glu Asn Arg Thr Pro His Leu Lys Asn Arg Asn Arg Asp
3445 3450 3455

aaa aac tac gga tgg aga acc gga ctc cac aca ttg aga cag aag aag 10534
Lys Asn Tyr Gly Trp Arg Thr Gly Leu His Thr Leu Arg Gln Lys Lys
3460 3465 3470

ttg tca gcc cag aac tcc aca cga gtt ttg cca ctg cta agc tgt gag 10582
Leu Ser Ala Gln Asn Ser Thr Arg Val Leu Pro Leu Leu Ser Cys Glu
3475 3480 3485

gca gtg cag gct ggg aca gcc gac ctc cag gtt gcg aaa aac ctg gtt 10630
Ala Val Gln Ala Gly Thr Ala Asp Leu Gln Val Ala Lys Asn Leu Val
3490 3495 3500

tct ggg acc tcc cac ccc aga gta aaa aga acg gag cct ccg cta cca 10678
Ser Gly Thr Ser His Pro Arg Val Lys Arg Thr Glu Pro Pro Leu Pro
3505 3510 3515 3520

ccc tcc cac gtg gtg gta gaa aga cgg ggt cta gag gtt aga gga gac 10726
Pro Ser His Val Val Val Glu Arg Arg Gly Leu Glu Val Arg Gly Asp
3525 3530 3535

cct cca ggg aac aaa tag tgggaccata ttgacgccag ggaaagaccg 10774
Pro Pro Gly Asn Lys
3540

gagtggttct ctgcttttcc tccaggggtc tgtgagcaca gtttgctcaa gaataagcag 10834

acctttggat gaaaaacaca aaaccact 10862

<210> 2
<211> 3411
<212> PRT
<213> Yellow fever virus

<400> 2
Met Ser Gly Arg Lys Ala Gln Gly Lys Thr Leu Gly Val Asn Met Val
1 5 10 15
Arg Arg Gly Val Arg Ser Leu Ser Asn Lys Ile Lys Gln Lys Thr Lys
20 25 30
Gln Ile Gly Asn Arg Pro Gly Pro Ser Arg Gly Val Gln Gly Phe Ile
35 40 45

Phe Phe Phe Leu Phe Asn Ile Leu Thr Gly Lys Lys Ile Thr Ala His
    50   55   60
Leu Lys Arg Leu Trp Lys Met Leu Asp Pro Arg Gln Gly Leu Ala Val
65   70   75   80
Leu Arg Lys Val Lys Arg Val Val Ala Ser Leu Met Arg Gly Leu Ser
                85   90   95
Ser Arg Lys Arg Arg Ser His Asp Val Leu Thr Val Gln Phe Leu Ile
            100   105   110
Leu Gly Met Leu Leu Met Thr Gly Gly Val Thr Leu Val Arg Lys Asn
        115   120   125
Arg Trp Leu Leu Leu Asn Val Thr Ser Glu Asp Leu Gly Lys Thr Phe
    130   135   140
Ser Val Gly Thr Gly Asn Cys Thr Thr Asn Ile Leu Glu Ala Lys Tyr
145   150   155   160
Trp Cys Pro Asp Ser Met Glu Tyr Asn Cys Pro Asn Leu Ser Pro Arg
                165   170   175
Glu Glu Pro Asp Asp Ile Asp Cys Trp Cys Tyr Gly Val Glu Asn Val
            180   185   190
Arg Val Ala Tyr Gly Lys Cys Asp Ser Ala Gly Arg Ser Arg Arg Ser
        195   200   205
Arg Arg Ala Ile Asp Leu Pro Thr His Glu Asn His Gly Leu Lys Thr
    210   215   220
Arg Gln Glu Lys Trp Met Thr Gly Arg Met Gly Glu Arg Gln Leu Gln
225   230   235   240
Lys Ile Glu Arg Trp Leu Val Arg Asn Pro Phe Phe Ala Val Thr Ala
                245   250   255
Leu Thr Ile Ala Tyr Leu Val Gly Ser Asn Met Thr Gln Arg Val Val
            260   265   270
Ile Ala Leu Leu Val Leu Ala Val Gly Pro Ala Tyr Ser Ala His Cys
        275   280   285
Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His Gly Gly Thr
    290   295   300
Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr Val Met Ala
305   310   315   320
Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val Ala Ile Asp
                325   330   335
Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val Leu Thr His
            340   345   350
Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala His Leu Ala
        355   360   365
Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr Ser Asp Arg
    370   375   380
Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser Ile Val Ala
385   390   395   400
Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe Glu Val Asp
                405   410   415
Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His Val Gly Ala
            420   425   430
Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys Phe Asp Ala
        435   440   445
Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly Lys Ala Thr
    450   455   460

Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn Ser Tyr Ile
465 470 475 480
Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln Trp Ala Gln
485 490 495
Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val Trp Arg Glu
500 505 510
Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala Thr Ile Arg
515 520 525
Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr Ala Leu Thr
530 535 540
Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn Leu Tyr Lys
545 550 555 560
Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser Ala Leu Thr
565 570 575
Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met Ser Phe Val
580 585 590
Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met Gln Val Arg
595 600 605
Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val Ala Asp Asp
610 615 620
Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val Asn Pro Ile
625 630 635 640
Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn Pro Pro Phe
645 650 655
Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg Leu Thr Tyr
660 665 670
Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe Thr Gln Thr
675 680 685
Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala Ala Trp Asp
690 695 700
Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys Gly Ile His
705 710 715 720
Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly Leu Asn Trp
725 730 735
Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val Gly Ile Asn
740 745 750
Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val Gly Val Ile
755 760 765
Met Met Phe Leu Ser Leu Gly Val Gly Ala Asp Gln Gly Cys Ala Ile
770 775 780
Asn Phe Gly Lys Arg Glu Leu Lys Cys Gly Asp Gly Ile Phe Ile Phe
785 790 795 800
Arg Asp Ser Asp Asp Trp Leu Asn Lys Tyr Ser Tyr Tyr Pro Glu Asp
805 810 815
Pro Val Lys Leu Ala Ser Ile Val Lys Ala Ser Phe Glu Glu Gly Lys
820 825 830
Cys Gly Leu Asn Ser Val Asp Ser Leu Glu His Glu Met Trp Arg Ser
835 840 845
Arg Ala Asp Glu Ile Asn Ala Ile Leu Glu Glu Asn Glu Val Asp Ile
850 855 860
Ser Val Val Val Gln Asp Pro Lys Asn Val Tyr Gln Arg Gly Thr His
865 870 875 880

Pro Phe Ser Arg Ile Arg Asp Gly Leu Gln Tyr Gly Trp Lys Thr Trp
885 890 895
Gly Lys Asn Leu Val Phe Ser Pro Gly Arg Lys Asn Gly Ser Phe Ile
900 905 910
Ile Asp Gly Lys Ser Arg Lys Glu Cys Pro Phe Ser Asn Arg Val Trp
915 920 925
Asn Ser Phe Gln Ile Glu Glu Phe Gly Thr Gly Val Phe Thr Thr Arg
930 935 940
Val Tyr Met Asp Ala Val Phe Glu Tyr Thr Ile Asp Cys Asp Gly Ser
945 950 955 960
Ile Leu Gly Ala Ala Val Asn Gly Lys Lys Ser Ala His Gly Ser Pro
965 970 975
Thr Phe Trp Met Gly Ser His Glu Val Asn Gly Thr Trp Met Ile His
980 985 990
Thr Leu Glu Ala Leu Asp Tyr Lys Glu Cys Glu Trp Pro Leu Thr His
995 1000 1005
Thr Ile Gly Thr Ser Val Glu Glu Ser Glu Met Phe Met Pro Arg Ser
1010 1015 1020
Ile Gly Gly Pro Val Ser Ser His Asn His Ile Pro Gly Tyr Lys Val
1025 1030 1035 1040
Gln Thr Asn Gly Pro Trp Met Gln Val Pro Leu Glu Val Lys Arg Glu
1045 1050 1055
Ala Cys Pro Gly Thr Ser Val Ile Ile Asp Gly Asn Cys Asp Gly Arg
1060 1065 1070
Gly Lys Ser Thr Arg Ser Thr Thr Asp Ser Gly Lys Ile Ile Pro Glu
1075 1080 1085
Trp Cys Cys Arg Ser Cys Thr Met Pro Pro Val Ser Phe His Gly Ser
1090 1095 1100
Asp Gly Cys Trp Tyr Pro Met Glu Ile Arg Pro Arg Lys Thr His Glu
1105 1110 1115 1120
Ser His Leu Val Arg Ser Trp Val Thr Ala Gly Glu Ile His Ala Val
1125 1130 1135
Pro Phe Gly Leu Val Ser Met Met Ile Ala Met Glu Val Val Leu Arg
1140 1145 1150
Lys Arg Gln Gly Pro Lys Gln Met Leu Val Gly Gly Val Val Leu Leu
1155 1160 1165
Gly Ala Met Leu Val Gly Gln Val Thr Leu Leu Asp Leu Leu Lys Leu
1170 1175 1180
Thr Val Ala Val Gly Leu His Phe His Glu Met Asn Asn Gly Gly Asp
1185 1190 1195 1200
Ala Met Tyr Met Ala Leu Ile Ala Ala Phe Ser Ile Arg Pro Gly Leu
1205 1210 1215
Leu Ile Gly Phe Gly Leu Arg Thr Leu Trp Ser Pro Arg Glu Arg Leu
1220 1225 1230
Val Leu Ala Leu Gly Ala Ala Met Val Glu Ile Ala Leu Gly Gly Met
1235 1240 1245
Met Gly Gly Leu Trp Lys Tyr Leu Asn Ala Val Ser Leu Cys Ile Leu
1250 1255 1260
Thr Ile Asn Ala Val Ala Ser Arg Lys Ala Ser Asn Thr Ile Leu Pro
1265 1270 1275 1280
Leu Met Ala Leu Leu Thr Pro Val Thr Met Ala Glu Val Arg Leu Ala
1285 1290 1295

Thr Met Leu Phe Cys Thr Val Val Ile Ile Gly Val Leu His Gln Asn
1300 1305 1310
Ser Lys Asp Thr Ser Met Gln Lys Thr Ile Pro Leu Val Ala Leu Thr
1315 1320 1325
Leu Thr Ser Tyr Leu Gly Leu Thr Gln Pro Phe Leu Gly Leu Cys Ala
1330 1335 1340
Phe Leu Ala Thr Arg Ile Phe Gly Arg Arg Ser Ile Pro Val Asn Glu
1345 1350 1355 1360
Ala Leu Ala Ala Ala Gly Leu Val Gly Val Leu Ala Gly Leu Ala Phe
1365 1370 1375
Gln Glu Met Glu Asn Phe Leu Gly Pro Ile Ala Val Gly Gly Ile Leu
1380 1385 1390
Met Met Leu Val Ser Val Ala Gly Arg Val Asp Gly Leu Glu Leu Lys
1395 1400 1405
Lys Leu Gly Glu Val Ser Trp Glu Glu Glu Ala Glu Ile Ser Gly Ser
1410 1415 1420
Ser Ala Arg Tyr Asp Val Ala Leu Ser Glu Gln Gly Glu Phe Lys Leu
1425 1430 1435 1440
Leu Ser Glu Glu Lys Val Pro Trp Asp Gln Val Val Met Thr Ser Leu
1445 1450 1455
Ala Leu Val Gly Ala Ala Ile His Pro Phe Ala Leu Leu Leu Val Leu
1460 1465 1470
Ala Gly Trp Leu Phe His Val Arg Gly Ala Arg Arg Ser Gly Asp Val
1475 1480 1485
Leu Trp Asp Ile Pro Thr Pro Lys Ile Ile Glu Glu Cys Glu His Leu
1490 1495 1500
Glu Asp Gly Ile Tyr Gly Ile Phe Gln Ser Thr Phe Leu Gly Ala Ser
1505 1510 1515 1520
Gln Arg Gly Val Gly Val Ala Gln Gly Gly Val Phe His Thr Met Trp
1525 1530 1535
His Val Thr Arg Gly Ala Phe Leu Val Arg Asn Gly Lys Lys Leu Ile
1540 1545 1550
Pro Ser Trp Ala Ser Val Lys Glu Asp Leu Val Ala Tyr Gly Gly Ser
1555 1560 1565
Trp Lys Leu Glu Gly Arg Trp Asp Gly Glu Glu Glu Val Gln Leu Ile
1570 1575 1580
Ala Ala Val Pro Gly Lys Asn Val Val Asn Val Gln Thr Lys Pro Ser
1585 1590 1595 1600
Leu Phe Lys Val Arg Asn Gly Gly Glu Ile Gly Ala Val Ala Leu Asp
1605 1610 1615
Tyr Pro Ser Gly Thr Ser Gly Ser Pro Ile Val Asn Arg Asn Gly Glu
1620 1625 1630
Val Ile Gly Leu Tyr Gly Asn Gly Ile Leu Val Gly Asp Asn Ser Phe
1635 1640 1645
Val Ser Ala Ile Ser Gln Thr Glu Val Lys Glu Glu Gly Lys Glu Glu
1650 1655 1660
Leu Gln Glu Ile Pro Thr Met Leu Lys Lys Gly Met Thr Thr Ile Leu
1665 1670 1675 1680
Asp Phe His Pro Gly Ala Gly Lys Thr Arg Arg Phe Leu Pro Gln Ile
1685 1690 1695
Leu Ala Glu Cys Ala Arg Arg Arg Leu Arg Thr Leu Val Leu Ala Pro
1700 1705 1710

Thr Arg Val Val Leu Ser Glu Met Lys Glu Ala Phe His Gly Leu Asp
1715 1720 1725
Val Lys Phe His Thr Gln Ala Phe Ser Ala His Gly Ser Gly Arg Glu
1730 1735 1740
Val Ile Asp Ala Met Cys His Ala Thr Leu Thr Tyr Arg Met Leu Glu
1745 1750 1755 1760
Pro Thr Arg Val Val Asn Trp Glu Val Ile Ile Met Asp Glu Ala His
1765 1770 1775
Phe Leu Asp Pro Ala Ser Ile Ala Ala Arg Gly Trp Ala Ala His Arg
1780 1785 1790
Ala Arg Ala Asn Glu Ser Ala Thr Ile Leu Met Thr Ala Thr Pro Pro
1795 1800 1805
Gly Thr Ser Asp Glu Phe Pro His Ser Asn Gly Glu Ile Glu Asp Val
1810 1815 1820
Gln Thr Asp Ile Pro Ser Glu Pro Trp Asn Thr Gly His Asp Trp Ile
1825 1830 1835 1840
Leu Ala Asp Lys Arg Pro Thr Ala Trp Phe Leu Pro Ser Ile Arg Ala
1845 1850 1855
Ala Asn Val Met Ala Ala Ser Leu Arg Lys Ala Gly Lys Ser Val Val
1860 1865 1870
Val Leu Asn Arg Lys Thr Phe Glu Arg Glu Tyr Pro Thr Ile Lys Gln
1875 1880 1885
Lys Lys Pro Asp Phe Ile Leu Ala Thr Asp Ile Ala Glu Met Gly Ala
1890 1895 1900
Asn Leu Cys Val Glu Arg Val Leu Asp Cys Arg Thr Ala Phe Lys Pro
1905 1910 1915 1920
Val Leu Val Asp Glu Gly Arg Lys Val Ala Ile Lys Gly Pro Leu Arg
1925 1930 1935
Ile Ser Ala Ser Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn
1940 1945 1950
Pro Asn Arg Asp Gly Asp Ser Tyr Tyr Tyr Ser Glu Pro Thr Ser Glu
1955 1960 1965
Asp Asn Ala His His Val Cys Trp Leu Glu Ala Ser Met Leu Leu Asp
1970 1975 1980
Asn Met Glu Val Arg Gly Gly Met Val Ala Pro Leu Tyr Gly Val Glu
1985 1990 1995 2000
Gly Thr Lys Thr Pro Val Ser Pro Gly Glu Met Arg Leu Arg Asp Asp
2005 2010 2015
Gln Arg Lys Val Phe Arg Glu Leu Val Arg Asn Cys Asp Leu Pro Val
2020 2025 2030
Trp Leu Ser Trp Gln Val Ala Lys Ala Gly Leu Lys Thr Asn Asp Arg
2035 2040 2045
Lys Trp Cys Phe Glu Gly Pro Glu Glu His Glu Ile Leu Asn Asp Ser
2050 2055 2060
Gly Glu Thr Val Lys Cys Arg Ala Pro Gly Gly Ala Lys Lys Pro Leu
2065 2070 2075 2080
Arg Pro Arg Trp Cys Asp Glu Arg Val Ser Ser Asp Gln Ser Ala Leu
2085 2090 2095
Ser Glu Phe Ile Lys Phe Ala Glu Gly Arg Arg Gly Ala Ala Glu Val
2100 2105 2110
Leu Val Val Leu Ser Glu Leu Pro Asp Phe Leu Ala Lys Lys Gly Gly
2115 2120 2125

Glu Ala Met Asp Thr Ile Ser Val Phe Leu His Ser Glu Glu Gly Ser
2130 2135 2140

Arg Ala Tyr Arg Asn Ala Leu Ser Met Met Pro Glu Ala Met Thr Ile
2145 2150 2155 2160

Val Met Leu Phe Ile Leu Ala Gly Leu Leu Thr Ser Gly Met Val Ile
2165 2170 2175

Phe Phe Met Ser Pro Lys Gly Ile Ser Arg Met Ser Met Ala Met Gly
2180 2185 2190

Thr Met Ala Gly Cys Gly Tyr Leu Met Phe Leu Gly Gly Val Lys Pro
2195 2200 2205

Thr His Ile Ser Tyr Ile Met Leu Ile Phe Phe Val Leu Met Val Val
2210 2215 2220

Val Ile Pro Glu Pro Gly Gln Gln Arg Ser Ile Gln Asp Asn Gln Val
2225 2230 2235 2240

Ala Tyr Leu Ile Ile Gly Ile Leu Thr Leu Val Ser Val Val Ala Ala
2245 2250 2255

Asn Glu Leu Gly Met Leu Glu Lys Thr Lys Glu Asp Leu Phe Gly Lys
2260 2265 2270

Lys Asn Leu Ile Pro Ser Ser Ala Ser Pro Trp Ser Trp Pro Asp Leu
2275 2280 2285

Asp Leu Lys Pro Gly Ala Ala Trp Thr Val Tyr Val Gly Ile Val Thr
2290 2295 2300

Met Leu Ser Pro Met Leu His His Trp Ile Lys Val Glu Tyr Gly Asn
2305 2310 2315 2320

Leu Ser Leu Ser Gly Ile Ala Gln Ser Ala Ser Val Leu Ser Phe Met
2325 2330 2335

Asp Lys Gly Ile Pro Phe Met Lys Met Asn Ile Ser Val Ile Ile Leu
2340 2345 2350

Leu Ile Ser Gly Trp Asn Ser Ile Thr Val Met Pro Leu Leu Cys Gly
2355 2360 2365

Ile Gly Cys Ala Met Leu His Trp Ser Leu Ile Leu Pro Gly Ile Lys
2370 2375 2380

Ala Gln Gln Ser Lys Leu Ala Gln Arg Arg Val Phe His Gly Val Ala
2385 2390 2395 2400

Lys Asn Pro Val Val Asp Gly Asn Pro Thr Val Asp Ile Glu Glu Ala
2405 2410 2415

Pro Glu Met Pro Ala Leu Tyr Glu Lys Lys Leu Ala Leu Tyr Leu Leu
2420 2425 2430

Leu Ala Leu Ser Leu Ala Ser Val Ala Met Cys Arg Thr Pro Phe Ser
2435 2440 2445

Leu Ala Glu Gly Ile Val Leu Ala Ser Ala Ala Leu Gly Pro Leu Ile
2450 2455 2460

Glu Gly Asn Thr Ser Leu Leu Trp Asn Gly Pro Met Ala Val Ser Met
2465 2470 2475 2480

Thr Gly Val Met Arg Gly Asn Tyr Tyr Ala Phe Val Gly Val Met Tyr
2485 2490 2495

Asn Leu Trp Lys Met Lys Thr Gly Arg Arg Gly Ser Ala Asn Gly Lys
2500 2505 2510

Thr Leu Gly Glu Val Trp Lys Arg Glu Leu Asn Leu Leu Asp Lys Gln
2515 2520 2525

Gln Phe Glu Leu Tyr Lys Arg Thr Asp Ile Val Glu Val Asp Arg Asp
2530 2535 2540

Thr Ala Arg Arg His Leu Ala Glu Gly Lys Val Asp Thr Gly Val Ala
2545 2550 2555 2560
Val Ser Arg Gly Thr Ala Lys Leu Arg Trp Phe His Glu Arg Gly Tyr
2565 2570 2575
Val Lys Leu Glu Gly Arg Val Ile Asp Leu Gly Cys Gly Arg Gly Gly
2580 2585 2590
Trp Cys Tyr Tyr Ala Ala Ala Gln Lys Glu Val Ser Gly Val Lys Gly
2595 2600 2605
Phe Thr Leu Gly Arg Asp Gly His Glu Lys Pro Met Asn Val Gln Ser
2610 2615 2620
Leu Gly Trp Asn Ile Ile Thr Phe Lys Asp Lys Thr Asp Ile His Arg
2625 2630 2635 2640
Leu Glu Pro Val Lys Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser
2645 2650 2655
Ser Ser Ser Ser Val Thr Glu Gly Glu Arg Thr Val Arg Val Leu Asp
2660 2665 2670
Thr Val Glu Lys Trp Leu Ala Cys Gly Val Asp Asn Phe Cys Val Lys
2675 2680 2685
Val Leu Ala Pro Tyr Met Pro Asp Val Leu Glu Lys Leu Glu Leu Leu
2690 2695 2700
Gln Arg Arg Phe Gly Gly Thr Val Ile Arg Asn Pro Leu Ser Arg Asn
2705 2710 2715 2720
Ser Thr His Glu Met Tyr Tyr Val Ser Gly Ala Arg Ser Asn Val Thr
2725 2730 2735
Phe Thr Val Asn Gln Thr Ser Arg Leu Leu Met Arg Arg Met Arg Arg
2740 2745 2750
Pro Thr Gly Lys Val Thr Leu Glu Ala Asp Val Ile Leu Pro Ile Gly
2755 2760 2765
Thr Arg Ser Val Glu Thr Asp Lys Gly Pro Leu Asp Lys Glu Ala Ile
2770 2775 2780
Glu Glu Arg Val Glu Arg Ile Lys Ser Glu Tyr Met Thr Ser Trp Phe
2785 2790 2795 2800
Tyr Asp Asn Asp Asn Pro Tyr Arg Thr Trp His Tyr Cys Gly Ser Tyr
2805 2810 2815
Val Thr Lys Thr Ser Gly Ser Ala Ala Ser Met Val Asn Gly Val Ile
2820 2825 2830
Lys Ile Leu Thr Tyr Pro Trp Asp Arg Ile Glu Glu Val Thr Arg Met
2835 2840 2845
Ala Met Thr Asp Thr Thr Pro Phe Gly Gln Gln Arg Val Phe Lys Glu
2850 2855 2860
Lys Val Asp Thr Arg Ala Lys Asp Pro Pro Ala Gly Thr Arg Lys Ile
2865 2870 2875 2880
Met Lys Val Val Asn Arg Trp Leu Phe Arg His Leu Ala Arg Glu Lys
2885 2890 2895
Asn Pro Arg Leu Cys Thr Lys Glu Glu Phe Ile Ala Lys Val Arg Ser
2900 2905 2910
His Ala Ala Ile Gly Ala Tyr Leu Glu Glu Gln Glu Gln Trp Lys Thr
2915 2920 2925
Ala Asn Glu Ala Val Gln Asp Pro Lys Phe Trp Glu Leu Val Asp Glu
2930 2935 2940
Glu Arg Lys Leu His Gln Gln Gly Arg Cys Arg Thr Cys Val Tyr Asn
2945 2950 2955 2960

Met Met Gly Lys Arg Glu Lys Lys Leu Ser Glu Phe Gly Lys Ala Lys
2965 2970 2975
Gly Ser Arg Ala Ile Trp Tyr Met Trp Leu Gly Ala Arg Tyr Leu Glu
2980 2985 2990
Phe Glu Ala Leu Gly Phe Leu Asn Glu Asp His Trp Ala Ser Arg Glu
2995 3000 3005
Asn Ser Gly Gly Gly Val Glu Gly Ile Gly Leu Gln Tyr Leu Gly Tyr
3010 3015 3020
Val Ile Arg Asp Leu Ala Ala Met Asp Gly Gly Gly Phe Tyr Ala Asp
3025 3030 3035 3040
Asp Thr Ala Gly Trp Asp Thr Arg Ile Thr Glu Ala Asp Leu Asp Asp
3045 3050 3055
Glu Gln Glu Ile Leu Asn Tyr Met Ser Pro His His Lys Lys Leu Ala
3060 3065 3070
Gln Ala Val Met Glu Met Thr Tyr Lys Asn Lys Val Val Lys Val Leu
3075 3080 3085
Arg Pro Ala Pro Gly Gly Lys Ala Tyr Met Asp Val Ile Ser Arg Arg
3090 3095 3100
Asp Gln Arg Gly Ser Gly Gln Val Val Thr Tyr Ala Leu Asn Thr Ile
3105 3110 3115 3120
Thr Asn Leu Lys Val Gln Leu Ile Arg Met Ala Glu Ala Glu Met Val
3125 3130 3135
Ile His His Gln His Val Gln Asp Cys Asp Glu Ser Val Leu Thr Arg
3140 3145 3150
Leu Glu Ala Trp Leu Thr Glu His Gly Cys Asn Arg Leu Lys Arg Met
3155 3160 3165
Ala Val Ser Gly Asp Asp Cys Val Val Arg Pro Ile Asp Asp Arg Phe
3170 3175 3180
Gly Leu Ala Leu Ser His Leu Asn Ala Met Ser Lys Val Arg Lys Asp
3185 3190 3195 3200
Ile Ser Glu Trp Gln Pro Ser Lys Gly Trp Asn Asp Trp Glu Asn Val
3205 3210 3215
Pro Phe Cys Ser His His Phe His Glu Leu Gln Leu Lys Asp Gly Arg
3220 3225 3230
Arg Ile Val Val Pro Cys Arg Glu Gln Asp Glu Leu Ile Gly Arg Gly
3235 3240 3245
Arg Val Ser Pro Gly Asn Gly Trp Met Ile Lys Glu Thr Ala Cys Leu
3250 3255 3260
Ser Lys Ala Tyr Ala Asn Met Trp Ser Leu Met Tyr Phe His Lys Arg
3265 3270 3275 3280
Asp Met Arg Leu Leu Ser Leu Ala Val Ser Ser Ala Val Pro Thr Ser
3285 3290 3295
Trp Val Pro Gln Gly Arg Thr Thr Trp Ser Ile His Gly Lys Gly Glu
3300 3305 3310
Trp Met Thr Thr Glu Asp Met Leu Glu Val Trp Asn Arg Val Trp Ile
3315 3320 3325
Thr Asn Asn Pro His Met Gln Asp Lys Thr Met Val Lys Glu Trp Arg
3330 3335 3340
Asp Val Pro Tyr Leu Thr Lys Arg Gln Asp Lys Leu Cys Gly Ser Leu
3345 3350 3355 3360
Ile Gly Met Thr Asn Arg Ala Thr Trp Ala Ser His Ile His Leu Val
3365 3370 3375

Ile His Arg Ile Arg Thr Leu Ile Gly Gln Glu Lys Tyr Thr Asp Tyr
3380 3385 3390
Leu Thr Val Met Asp Arg Tyr Ser Val Asp Ala Asp Leu Gln Pro Gly
3395 3400 3405
Glu Leu Ile
3410

<210> 3
<211> 1479
<212> DNA
<213> Yellow fever virus

<220>
<221> CDS
<222> (1)..(1479)

<400> 3
gct cac tgc att gga att act gac aga gat ttc att gag ggg gtg cat 48
Ala His Cys Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His
1 5 10 15

gga gga act tgg gtt tca gct acc ctg gag cac ggc aag tgt gtc act 96
Gly Gly Thr Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr
20 25 30

gtt atg gcc cct gac aag cct tca ttg gac atc tca cta gag aca gta 144
Val Met Ala Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val
35 40 45

gcc att gat gga cct gct gag gcg agg aaa gtg tgt tac aat gca gtt 192
Ala Ile Asp Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val
50 55 60

ctc act cat gtg aag att aat gac aag tgc ccc agc act gga gag gcc 240
Leu Thr His Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala
65 70 75 80

cac cta gct gaa gag aac gaa ggg gac aat gcg tgc aag cgc act tat 288
His Leu Ala Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr
85 90 95

tct gat aga ggc tgg ggc aat ggc tgt ggc cta ttt ggg aaa ggg agc 336
Ser Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser
100 105 110

att gtg gca tgc gcc aaa ttc act tgt gcc aaa tcc atg agt ttg ttt 384
Ile Val Ala Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe
115 120 125

```
gag gtt gat cag acc aaa att cag tat gtc atc aga gca caa ttg cat    432
Glu Val Asp Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His
    130                 135                 140

gta ggg gcc aag cag gaa aat tgg aat acc gcc att aag act ctc aag    480
Val Gly Ala Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys
145                 150                 155                 160

ttt gat gcc ctg tca ggc tcc cag gaa gcc gag ttc act ggg tat gga    528
Phe Asp Ala Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly
                165                 170                 175

aaa gct aca ctg gaa tgc cag gtg caa act gcg gtg gac ttt ggt aac    576
Lys Ala Thr Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn
            180                 185                 190

agt tac atc gct gag atg gaa aaa gag agc tgg ata gtg gac aga cag    624
Ser Tyr Ile Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln
        195                 200                 205

tgg gcc cag gac ttg acc ctg cca tgg cag agt gga agt ggc ggg gtg    672
Trp Ala Gln Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val
    210                 215                 220

tgg aga gag atg cat cat ctt gtc gaa ttt gaa cct ccg cat gcc gcc    720
Trp Arg Glu Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala
225                 230                 235                 240

act atc aga gta ctg gcc ctg gga aac cag gaa ggc tcc ttg aaa aca    768
Thr Ile Arg Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr
                245                 250                 255

gct ctt acc ggc gca atg agg gtt aca aag gac aca aat gac aac aac    816
Ala Leu Thr Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn
            260                 265                 270

ctt tac aaa cta cat ggt gga cat gtt tcc tgc aga gtg aaa ttg tca    864
Leu Tyr Lys Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser
        275                 280                 285

gct ttg aca ctc aag ggg aca tcc tac aaa atg tgc act gac aaa atg    912
Ala Leu Thr Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met
    290                 295                 300

tct ttt gtc aag aac cca act gac act ggc cat ggc act gtt gtg atg    960
Ser Phe Val Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met
305                 310                 315                 320

cag gtg aga gtg cca aaa gga gcc ccc tgc agg att cca gtg ata gta    1008
Gln Val Arg Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val
                325                 330                 335
```

gct gat gat ctt aca gcg gca atc aat aaa ggc att ttg gtt aca gtt 1056
Ala Asp Asp Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val
340 345 350

aac ccc atc gcc tca acc aat gat gat gaa gtg ctg att gag gtg aac 1104
Asn Pro Ile Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn
355 360 365

cca cct ttt gga gac agc tac att atc gtt ggg aca gga gat tca cgt 1152
Pro Pro Phe Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg
370 375 380

ctc act tac cag tgg cac aaa gag gga agc tca ata gga aag ttg ttc 1200
Leu Thr Tyr Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe
385 390 395 400

act cag acc atg aaa ggc gcg gaa cgc ctg gcc gtc atg gga gac gcc 1248
Thr Gln Thr Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala
405 410 415

gcc tgg gat ttc agc tcc gct gga ggg ttc ttc act tcg gtt ggg aaa 1296
Ala Trp Asp Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys
420 425 430

gga att cat acg gtg ttt ggc tct gcc ttt cag ggg cta ttt ggc ggc 1344
Gly Ile His Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly
435 440 445

ttg aac tgg ata aca aag gtc atc atg ggg gcg gta ctc ata tgg gtt 1392
Leu Asn Trp Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val
450 455 460

ggc atc aac aca aga aac atg aca atg tcc atg agc atg atc ttg gta 1440
Gly Ile Asn Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val
465 470 475 480

gga gtg atc atg atg ttt ttg tct cta gga gtt ggg gcg 1479
Gly Val Ile Met Met Phe Leu Ser Leu Gly Val Gly Ala
485 490

<210> 4
<211> 493
<212> PRT
<213> Yellow fever virus

<400> 4
Ala His Cys Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His
1 5 10 15

Gly Gly Thr Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr

20 25 30

Val Met Ala Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val
35 40 45

Ala Ile Asp Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val
50 55 60

Leu Thr His Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala
65 70 75 80

His Leu Ala Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr
85 90 95

Ser Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser
100 105 110

Ile Val Ala Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe
115 120 125

Glu Val Asp Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His
130 135 140

Val Gly Ala Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys
145 150 155 160

Phe Asp Ala Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly
165 170 175

Lys Ala Thr Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn
180 185 190

Ser Tyr Ile Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln
195 200 205

Trp Ala Gln Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val
210 215 220

Trp Arg Glu Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala
225 230 235 240

Thr Ile Arg Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr
245 250 255

Ala Leu Thr Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn
260 265 270

Leu Tyr Lys Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser
275 280 285

Ala Leu Thr Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met
290 295 300

Ser Phe Val Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met
305 310 315 320

Gln Val Arg Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val
325 330 335

Ala Asp Asp Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val
340 345 350

Asn Pro Ile Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn
355 360 365

Pro Pro Phe Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg
370 375 380

Leu Thr Tyr Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe
385 390 395 400

Thr Gln Thr Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala
405 410 415

Ala Trp Asp Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys
420 425 430

Gly Ile His Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly
435 440 445

Leu Asn Trp Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val
450 455 460

Gly Ile Asn Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val
465 470 475 480

Gly Val Ile Met Met Phe Leu Ser Leu Gly Val Gly Ala
485 490

SEQUENCE LISTING

<110> BARRETT, ALAN
MCARTHUR, MONICA

<120> METHODS AND COMPOSITIONS CONCERNING ALTERED YELLOW
FEVER VIRUS STRAINS

<130> UTSG:255USP1

<140> UNKNOWN
<141> 2002-07-19

<160> 4

<170> PatentIn Ver. 2.1

<210> 1
<211> 10862
<212> DNA
<213> Yellow fever virus

<220>
<221> CDS
<222> (119)..(10744)

<400> 1
```
agtaaatcct gtgtgctaat tgaggtgcat tggtctgcaa atcgagttgc taggcaataa 60

acacatttgg attaatttta atcgttcgtt gagcgattag cagagaactg accagaac   118

atg tct ggt cgt aaa gct cag gga aaa acc ctg ggc gtc aat atg gta   166
Met Ser Gly Arg Lys Ala Gln Gly Lys Thr Leu Gly Val Asn Met Val
  1               5                  10                  15

cga cga gga gtt cgc tcc ttg tca aac aaa ata aaa caa aaa aca aaa   214
Arg Arg Gly Val Arg Ser Leu Ser Asn Lys Ile Lys Gln Lys Thr Lys
             20                  25                  30

caa att gga aac aga cct gga cct tca aga ggt gtt caa gga ttt atc   262
Gln Ile Gly Asn Arg Pro Gly Pro Ser Arg Gly Val Gln Gly Phe Ile
         35                  40                  45

ttt ttc ttt ttg ttc aac att ttg act gga aaa aag atc acg gcc cac   310
Phe Phe Phe Leu Phe Asn Ile Leu Thr Gly Lys Lys Ile Thr Ala His
     50                  55                  60

cta aag agg ttg tgg aaa atg ctg gac cca aga caa ggc ttg gct gtt   358
Leu Lys Arg Leu Trp Lys Met Leu Asp Pro Arg Gln Gly Leu Ala Val
 65                  70                  75                  80

cta agg aaa gtt aag aga gtg gtg gcc agt ttg atg aga gga ttg tcc   406
Leu Arg Lys Val Lys Arg Val Val Ala Ser Leu Met Arg Gly Leu Ser
                 85                  90                  95
```

```
tca agg aaa cgc cgt tcc cat gat gtt ctg act gtg caa ttc cta att    454
Ser Arg Lys Arg Arg Ser His Asp Val Leu Thr Val Gln Phe Leu Ile
            100                 105                 110

ttg gga atg ctg ttg atg acg ggt gga gtg acc ttg gtg cgg aaa aac    502
Leu Gly Met Leu Leu Met Thr Gly Gly Val Thr Leu Val Arg Lys Asn
        115                 120                 125

aga tgg ttg ctc cta aat gtg aca tct gag gac ctc ggg aaa aca ttc    550
Arg Trp Leu Leu Leu Asn Val Thr Ser Glu Asp Leu Gly Lys Thr Phe
    130                 135                 140

tct gtg ggc aca ggc aac tgc aca aca aac att ttg gaa gcc aag tac    598
Ser Val Gly Thr Gly Asn Cys Thr Thr Asn Ile Leu Glu Ala Lys Tyr
145                 150                 155                 160

tgg tgc cca gac tca atg gaa tac aac tgt ccc aat ctc agt cca aga    646
Trp Cys Pro Asp Ser Met Glu Tyr Asn Cys Pro Asn Leu Ser Pro Arg
                165                 170                 175

gag gag cca gat gac att gat tgc tgg tgc tat ggg gtg gaa aac gtt    694
Glu Glu Pro Asp Asp Ile Asp Cys Trp Cys Tyr Gly Val Glu Asn Val
            180                 185                 190

aga gtc gca tat ggt aag tgt gac tca gca ggc agg tct agg agg tca    742
Arg Val Ala Tyr Gly Lys Cys Asp Ser Ala Gly Arg Ser Arg Arg Ser
        195                 200                 205

aga agg gcc att gac ttg cct acg cat gaa aac cat ggt ttg aag acc    790
Arg Arg Ala Ile Asp Leu Pro Thr His Glu Asn His Gly Leu Lys Thr
    210                 215                 220

cgg caa gaa aag tgg atg act gga aga atg ggt gaa agg caa ctc caa    838
Arg Gln Glu Lys Trp Met Thr Gly Arg Met Gly Glu Arg Gln Leu Gln
225                 230                 235                 240

aag att gag aga tgg ctc gtg agg aac ccc ttt ttt gca gtg aca gct    886
Lys Ile Glu Arg Trp Leu Val Arg Asn Pro Phe Phe Ala Val Thr Ala
                245                 250                 255

ttg acc att gcc tac ctt gtg gga agc aac atg acg caa cga gtc gtg    934
Leu Thr Ile Ala Tyr Leu Val Gly Ser Asn Met Thr Gln Arg Val Val
            260                 265                 270

att gcc cta ctg gtc ttg gct gtt ggt ccg gcc tac tca gct cac tgc    982
Ile Ala Leu Leu Val Leu Ala Val Gly Pro Ala Tyr Ser Ala His Cys
        275                 280                 285

att gga att act gac aga gat ttc att gag ggg gtg cat gga gga act    1030
Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His Gly Gly Thr
    290                 295                 300

tgg gtt tca gct acc ctg gag cac ggc aag tgt gtc act gtt atg gcc    1078
Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr Val Met Ala
305                 310                 315                 320
```

cct gac aag cct tca ttg gac atc tca cta gag aca gta gcc att gat 1126
Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val Ala Ile Asp
325 330 335

gga cct gct gag gcg agg aaa gtg tgt tac aat gca gtt ctc act cat 1174
Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val Leu Thr His
340 345 350

gtg aag att aat gac aag tgc ccc agc act gga gag gcc cac cta gct 1222
Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala His Leu Ala
355 360 365

gaa gag aac gaa ggg gac aat gcg tgc aag cgc act tat tct gat aga 1270
Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr Ser Asp Arg
370 375 380

ggc tgg ggc aat ggc tgt ggc cta ttt ggg aaa ggg agc att gtg gca 1318
Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser Ile Val Ala
385 390 395 400

tgc gcc aaa ttc act tgt gcc aaa tcc atg agt ttg ttt gag gtt gat 1366
Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe Glu Val Asp
405 410 415

cag acc aaa att cag tat gtc atc aga gca caa ttg cat gta ggg gcc 1414
Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His Val Gly Ala
420 425 430

aag cag gaa aat tgg aat acc gcc att aag act ctc aag ttt gat gcc 1462
Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys Phe Asp Ala
435 440 445

ctg tca ggc tcc cag gaa gcc gag ttc act ggg tat gga aaa gct aca 1510
Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly Lys Ala Thr
450 455 460

ctg gaa tgc cag gtg caa act gcg gtg gac ttt ggt aac agt tac atc 1558
Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn Ser Tyr Ile
465 470 475 480

gct gag atg gaa aaa gag agc tgg ata gtg gac aga cag tgg gcc cag 1606
Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln Trp Ala Gln
485 490 495

gac ttg acc ctg cca tgg cag agt gga agt ggc ggg gtg tgg aga gag 1654
Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val Trp Arg Glu
500 505 510

atg cat cat ctt gtc gaa ttt gaa cct ccg cat gcc gcc act atc aga 1702
Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala Thr Ile Arg
515 520 525

gta ctg gcc ctg gga aac cag gaa ggc tcc ttg aaa aca gct ctt acc 1750
Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr Ala Leu Thr

530 535 540

ggc gca atg agg gtt aca aag gac aca aat gac aac aac ctt tac aaa 1798
Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn Leu Tyr Lys
545 550 555 560

cta cat ggt gga cat gtt tcc tgc aga gtg aaa ttg tca gct ttg aca 1846
Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser Ala Leu Thr
565 570 575

ctc aag ggg aca tcc tac aaa atg tgc act gac aaa atg tct ttt gtc 1894
Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met Ser Phe Val
580 585 590

aag aac cca act gac act ggc cat ggc act gtt gtg atg cag gtg aga 1942
Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met Gln Val Arg
595 600 605

gtg cca aaa gga gcc ccc tgc agg att cca gtg ata gta gct gat gat 1990
Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val Ala Asp Asp
610 615 620

ctt aca gcg gca atc aat aaa ggc att ttg gtt aca gtt aac ccc atc 2038
Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val Asn Pro Ile
625 630 635 640

gcc tca acc aat gat gat gaa gtg ctg att gag gtg aac cca cct ttt 2086
Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn Pro Pro Phe
645 650 655

gga gac agc tac att atc gtt ggg aca gga gat tca cgt ctc act tac 2134
Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg Leu Thr Tyr
660 665 670

cag tgg cac aaa gag gga agc tca ata gga aag ttg ttc act cag acc 2182
Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe Thr Gln Thr
675 680 685

atg aaa ggc gcg gaa cgc ctg gcc gtc atg gga gac gcc gcc tgg gat 2230
Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala Ala Trp Asp
690 695 700

ttc agc tcc gct gga ggg ttc ttc act tcg gtt ggg aaa gga att cat 2278
Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys Gly Ile His
705 710 715 720

acg gtg ttt ggc tct gcc ttt cag ggg cta ttt ggc ggc ttg aac tgg 2326
Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly Leu Asn Trp
725 730 735

ata aca aag gtc atc atg ggg gcg gta ctc ata tgg gtt ggc atc aac 2374
Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val Gly Ile Asn
740 745 750

aca aga aac atg aca atg tcc atg agc atg atc ttg gta gga gtg atc 2422

Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val Gly Val Ile
755 760 765

atg atg ttt ttg tct cta gga gtt ggg gcg gat caa gga tgc gcc atc 2470
Met Met Phe Leu Ser Leu Gly Val Gly Ala Asp Gln Gly Cys Ala Ile
770 775 780

aac ttt ggc aag aga gag ctc aag tgc gga gat ggt atc ttc ata ttt 2518
Asn Phe Gly Lys Arg Glu Leu Lys Cys Gly Asp Gly Ile Phe Ile Phe
785 790 795 800

aga gac tct gat gac tgg ctg aac aag tac tca tac tat cca gaa gat 2566
Arg Asp Ser Asp Asp Trp Leu Asn Lys Tyr Ser Tyr Tyr Pro Glu Asp
805 810 815

cct gtg aag ctt gca tca ata gtg aaa gcc tct ttt gaa gaa ggg aag 2614
Pro Val Lys Leu Ala Ser Ile Val Lys Ala Ser Phe Glu Glu Gly Lys
820 825 830

tgt ggc cta aat tca gtt gac tcc ctt gag cat gag atg tgg aga agc 2662
Cys Gly Leu Asn Ser Val Asp Ser Leu Glu His Glu Met Trp Arg Ser
835 840 845

agg gca gat gag atc aat gcc att ctt gag gaa aac gag gtg gac att 2710
Arg Ala Asp Glu Ile Asn Ala Ile Leu Glu Glu Asn Glu Val Asp Ile
850 855 860

tct gtt gtc gtg cag gat cca aag aat gtt tac cag aga gga act cat 2758
Ser Val Val Val Gln Asp Pro Lys Asn Val Tyr Gln Arg Gly Thr His
865 870 875 880

cca ttt tcc aga att cgg gac ggt ctg cag tat ggt tgg aag act tgg 2806
Pro Phe Ser Arg Ile Arg Asp Gly Leu Gln Tyr Gly Trp Lys Thr Trp
885 890 895

ggt aag aac ctt gtg ttc tcc cca ggg agg aag aat gga agc ttc atc 2854
Gly Lys Asn Leu Val Phe Ser Pro Gly Arg Lys Asn Gly Ser Phe Ile
900 905 910

ata gat gga aag tcc agg aaa gaa tgc ccg ttt tca aac cgg gtc tgg 2902
Ile Asp Gly Lys Ser Arg Lys Glu Cys Pro Phe Ser Asn Arg Val Trp
915 920 925

aat tct ttc cag ata gag gag ttt ggg acg gga gtg ttc acc aca cgc 2950
Asn Ser Phe Gln Ile Glu Glu Phe Gly Thr Gly Val Phe Thr Thr Arg
930 935 940

gtg tac atg gac gca gtc ttt gaa tac acc ata gac tgc gat gga tct 2998
Val Tyr Met Asp Ala Val Phe Glu Tyr Thr Ile Asp Cys Asp Gly Ser
945 950 955 960

atc ttg ggt gca gcg gtg aac gga aaa aag agt gcc cat ggc tct cca 3046
Ile Leu Gly Ala Ala Val Asn Gly Lys Lys Ser Ala His Gly Ser Pro
965 970 975

```
aca ttt tgg atg gga agt cat gaa gta aat ggg aca tgg atg atc cac    3094
Thr Phe Trp Met Gly Ser His Glu Val Asn Gly Thr Trp Met Ile His
            980                 985                 990

acc ttg gag gca tta gat tac aag gag tgt gag tgg cca ctg aca cat    3142
Thr Leu Glu Ala Leu Asp Tyr Lys Glu Cys Glu Trp Pro Leu Thr His
        995                 1000                1005

acg att gga aca tca gtt gaa gag agt gaa atg ttc atg ccg aga tca    3190
Thr Ile Gly Thr Ser Val Glu Glu Ser Glu Met Phe Met Pro Arg Ser
    1010                1015                1020

atc gga ggc cca gtt agc tct cac aat cat atc cct gga tac aag gtt    3238
Ile Gly Gly Pro Val Ser Ser His Asn His Ile Pro Gly Tyr Lys Val
1025                1030                1035                1040

cag acg aac gga cct tgg atg cag gta cca cta gaa gtg aag aga gaa    3286
Gln Thr Asn Gly Pro Trp Met Gln Val Pro Leu Glu Val Lys Arg Glu
                1045                1050                1055

gct tgc cca ggg act agc gtg atc att gat ggc aac tgt gat gga cgg    3334
Ala Cys Pro Gly Thr Ser Val Ile Ile Asp Gly Asn Cys Asp Gly Arg
            1060                1065                1070

gga aaa tca acc aga tcc acc acg gat agc ggg aaa att att cct gaa    3382
Gly Lys Ser Thr Arg Ser Thr Thr Asp Ser Gly Lys Ile Ile Pro Glu
        1075                1080                1085

tgg tgt tgc cgc tcc tgc aca atg ccg cct gtg agc ttc cat ggt agt    3430
Trp Cys Cys Arg Ser Cys Thr Met Pro Pro Val Ser Phe His Gly Ser
    1090                1095                1100

gat ggg tgt tgg tat ccc atg gaa att agg cca agg aaa acg cat gaa    3478
Asp Gly Cys Trp Tyr Pro Met Glu Ile Arg Pro Arg Lys Thr His Glu
1105                1110                1115                1120

agc cat ctg gtg cgc tcc tgg gtt aca gct gga gaa ata cat gct gtc    3526
Ser His Leu Val Arg Ser Trp Val Thr Ala Gly Glu Ile His Ala Val
                1125                1130                1135

cct ttt ggt ttg gtg agc atg atg ata gca atg gaa gtg gtc cta agg    3574
Pro Phe Gly Leu Val Ser Met Met Ile Ala Met Glu Val Val Leu Arg
            1140                1145                1150

aaa aga cag gga cca aag caa atg ttg gtt gga gga gtg gtg ctc ttg    3622
Lys Arg Gln Gly Pro Lys Gln Met Leu Val Gly Gly Val Val Leu Leu
        1155                1160                1165

gga gca atg ctg gtc ggg caa gta act ctc ctt gat ttg ctg aaa ctc    3670
Gly Ala Met Leu Val Gly Gln Val Thr Leu Leu Asp Leu Leu Lys Leu
    1170                1175                1180

aca gtg gct gtg gga ttg cat ttc cat gag atg aac aat gga gga gac    3718
Thr Val Ala Val Gly Leu His Phe His Glu Met Asn Asn Gly Gly Asp
1185                1190                1195                1200
```

gcc atg tat atg gcg ttg att gct gcc ttt tca atc aga cca ggg ctg 3766
Ala Met Tyr Met Ala Leu Ile Ala Ala Phe Ser Ile Arg Pro Gly Leu
1205 1210 1215

ctc atc ggc ttt ggg ctc agg acc cta tgg agc cct cgg gaa cgc ctt 3814
Leu Ile Gly Phe Gly Leu Arg Thr Leu Trp Ser Pro Arg Glu Arg Leu
1220 1225 1230

gta ctg gcc cta gga gca gcc atg gtg gag att gcc ttg ggt ggc atg 3862
Val Leu Ala Leu Gly Ala Ala Met Val Glu Ile Ala Leu Gly Gly Met
1235 1240 1245

atg ggc ggc ctg tgg aag tat cta aat gca gtt tct ctc tgc atc ctg 3910
Met Gly Gly Leu Trp Lys Tyr Leu Asn Ala Val Ser Leu Cys Ile Leu
1250 1255 1260

aca ata aat gct gta gct tct agg aaa gca tca aat acc atc ttg ccc 3958
Thr Ile Asn Ala Val Ala Ser Arg Lys Ala Ser Asn Thr Ile Leu Pro
1265 1270 1275 1280

ctc atg gct ctg ttg aca cct gtc act atg gct gag gtg aga ctt gcc 4006
Leu Met Ala Leu Leu Thr Pro Val Thr Met Ala Glu Val Arg Leu Ala
1285 1290 1295

aca atg ctc ttt tgt acc gtg gtt atc ata ggg gtc ctt cac cag aac 4054
Thr Met Leu Phe Cys Thr Val Val Ile Ile Gly Val Leu His Gln Asn
1300 1305 1310

tcc aag gac acc tcc atg cag aag act ata cct ctg gtg gcc ctc aca 4102
Ser Lys Asp Thr Ser Met Gln Lys Thr Ile Pro Leu Val Ala Leu Thr
1315 1320 1325

ctc aca tct tac ctg ggc ttg aca caa cct ttt ttg ggc ctg tgt gca 4150
Leu Thr Ser Tyr Leu Gly Leu Thr Gln Pro Phe Leu Gly Leu Cys Ala
1330 1335 1340

ttt ctg gca acc cgc ata ttt ggg cga agg agt atc cca gtg aat gag 4198
Phe Leu Ala Thr Arg Ile Phe Gly Arg Arg Ser Ile Pro Val Asn Glu
1345 1350 1355 1360

gca ctc gca gca gct ggt cta gtg gga gtg ctg gca gga ctg gct ttt 4246
Ala Leu Ala Ala Ala Gly Leu Val Gly Val Leu Ala Gly Leu Ala Phe
1365 1370 1375

cag gag atg gag aac ttc ctt ggt ccg att gca gtt gga gga atc ctg 4294
Gln Glu Met Glu Asn Phe Leu Gly Pro Ile Ala Val Gly Gly Ile Leu
1380 1385 1390

atg atg ctg gtt agc gtg gct ggg agg gtg gat ggg cta gag ctc aag 4342
Met Met Leu Val Ser Val Ala Gly Arg Val Asp Gly Leu Glu Leu Lys
1395 1400 1405

aag ctt ggt gaa gtt tca tgg gaa gag gag gcg gag atc agc gga agt 4390
Lys Leu Gly Glu Val Ser Trp Glu Glu Glu Ala Glu Ile Ser Gly Ser

1410 1415 1420

tcc gcc cgc tat gat gtg gca ctc agt gaa caa ggg gag ttc aag ctg 4438
Ser Ala Arg Tyr Asp Val Ala Leu Ser Glu Gln Gly Glu Phe Lys Leu
1425 1430 1435 1440

ctt tct gaa gag aaa gtg cca tgg gac cag gtt gtg atg acc tcg ctg 4486
Leu Ser Glu Glu Lys Val Pro Trp Asp Gln Val Val Met Thr Ser Leu
1445 1450 1455

gcc ttg gtt ggg gct gcc att cat cca ttt gct ctt ctg ctg gtc ctt 4534
Ala Leu Val Gly Ala Ala Ile His Pro Phe Ala Leu Leu Leu Val Leu
1460 1465 1470

gct ggg tgg ctg ttt cat gtc agg gga gct agg aga agt ggg gat gtc 4582
Ala Gly Trp Leu Phe His Val Arg Gly Ala Arg Arg Ser Gly Asp Val
1475 1480 1485

ttg tgg gat att ccc act cct aag atc att gag gaa tgt gaa cat ctg 4630
Leu Trp Asp Ile Pro Thr Pro Lys Ile Ile Glu Glu Cys Glu His Leu
1490 1495 1500

gag gat ggg att tat ggc ata ttc cag tca acc ttc ttg ggg gcc tcc 4678
Glu Asp Gly Ile Tyr Gly Ile Phe Gln Ser Thr Phe Leu Gly Ala Ser
1505 1510 1515 1520

cag cga gga gtg gga gtg gca cag gga ggg gtg ttc cac aca atg tgg 4726
Gln Arg Gly Val Gly Val Ala Gln Gly Gly Val Phe His Thr Met Trp
1525 1530 1535

cat gtc aca aga gga gct ttc ctt gtc agg aat ggc aag aag ttg att 4774
His Val Thr Arg Gly Ala Phe Leu Val Arg Asn Gly Lys Lys Leu Ile
1540 1545 1550

cca tct tgg gct tca gta aag gaa gac ctt gtc gcc tat ggt ggc tca 4822
Pro Ser Trp Ala Ser Val Lys Glu Asp Leu Val Ala Tyr Gly Gly Ser
1555 1560 1565

tgg aag ttg gaa ggc aga tgg gat gga gag gaa gag gtc caa ttg atc 4870
Trp Lys Leu Glu Gly Arg Trp Asp Gly Glu Glu Glu Val Gln Leu Ile
1570 1575 1580

gct gct gtt cca gga aag aac gtg gtc aac gtc cag aca aaa ccg agc 4918
Ala Ala Val Pro Gly Lys Asn Val Val Asn Val Gln Thr Lys Pro Ser
1585 1590 1595 1600

ttg ttc aaa gtg agg aat ggg gga gaa atc ggg gct gtc gct ctt gac 4966
Leu Phe Lys Val Arg Asn Gly Gly Glu Ile Gly Ala Val Ala Leu Asp
1605 1610 1615

tat ccg agt ggc act tca gga tct cct att gtt aac agg aac gga gag 5014
Tyr Pro Ser Gly Thr Ser Gly Ser Pro Ile Val Asn Arg Asn Gly Glu
1620 1625 1630

gtg att ggg ctg tac ggc aat ggc atc ctt gtc ggt gac aac tcc ttc 5062

```
Val Ile Gly Leu Tyr Gly Asn Gly Ile Leu Val Gly Asp Asn Ser Phe
        1635                1640                1645

gtg tcc gcc ata tcc cag act gag gtg aag gaa gaa gga aag gag gag    5110
Val Ser Ala Ile Ser Gln Thr Glu Val Lys Glu Glu Gly Lys Glu Glu
    1650                1655                1660

ctc caa gag atc ccg aca atg cta aag aaa gga atg aca act atc ctt    5158
Leu Gln Glu Ile Pro Thr Met Leu Lys Lys Gly Met Thr Thr Ile Leu
1665                1670                1675                1680

gat ttt cat cct gga gct ggg aag aca aga cgt ttt ctc cca cag atc    5206
Asp Phe His Pro Gly Ala Gly Lys Thr Arg Arg Phe Leu Pro Gln Ile
                1685                1690                1695

ttg gcc gag tgc gca cgg aga cgc ttg cgc act ctt gtg ttg gcc ccc    5254
Leu Ala Glu Cys Ala Arg Arg Arg Leu Arg Thr Leu Val Leu Ala Pro
            1700                1705                1710

acc agg gtt gtt ctt tct gaa atg aag gag gct ttt cac ggc ctg gac    5302
Thr Arg Val Val Leu Ser Glu Met Lys Glu Ala Phe His Gly Leu Asp
        1715                1720                1725

gtg aaa ttc cac aca cag gct ttt tcc gct cac ggc agc ggg aga gaa    5350
Val Lys Phe His Thr Gln Ala Phe Ser Ala His Gly Ser Gly Arg Glu
    1730                1735                1740

gtc att gat gcc atg tgc cat gcc acc cta act tac agg atg ttg gaa    5398
Val Ile Asp Ala Met Cys His Ala Thr Leu Thr Tyr Arg Met Leu Glu
1745                1750                1755                1760

cca act agg gtt gtt aac tgg gaa gtg atc atc atg gat gaa gcc cat    5446
Pro Thr Arg Val Val Asn Trp Glu Val Ile Ile Met Asp Glu Ala His
                1765                1770                1775

ttt ttg gat cca gct agc ata gcc gcc aga ggt tgg gca gcg cac aga    5494
Phe Leu Asp Pro Ala Ser Ile Ala Ala Arg Gly Trp Ala Ala His Arg
            1780                1785                1790

gct agg gca aat gaa agt gca aca atc ttg atg aca gcc aca ccg cct    5542
Ala Arg Ala Asn Glu Ser Ala Thr Ile Leu Met Thr Ala Thr Pro Pro
        1795                1800                1805

ggg act agt gat gaa ttt cca cat tca aat ggt gaa ata gaa gat gtt    5590
Gly Thr Ser Asp Glu Phe Pro His Ser Asn Gly Glu Ile Glu Asp Val
    1810                1815                1820

caa acg gac ata ccc agt gag ccc tgg aac aca ggg cat gac tgg atc    5638
Gln Thr Asp Ile Pro Ser Glu Pro Trp Asn Thr Gly His Asp Trp Ile
1825                1830                1835                1840

ctg gct gac aaa agg ccc acg gca tgg ttc ctt cca tcc atc aga gct    5686
Leu Ala Asp Lys Arg Pro Thr Ala Trp Phe Leu Pro Ser Ile Arg Ala
                1845                1850                1855
```

gca aat gtc atg gct gcc tct ttg cgt aag gct gga aag agt gtg gtg 5734
Ala Asn Val Met Ala Ala Ser Leu Arg Lys Ala Gly Lys Ser Val Val
1860 1865 1870

gtc ctg aac agg aaa acc ttt gag aga gaa tac ccc acg ata aag cag 5782
Val Leu Asn Arg Lys Thr Phe Glu Arg Glu Tyr Pro Thr Ile Lys Gln
1875 1880 1885

aag aaa cct gac ttt ata ttg gcc act gac ata gct gaa atg gga gcc 5830
Lys Lys Pro Asp Phe Ile Leu Ala Thr Asp Ile Ala Glu Met Gly Ala
1890 1895 1900

aac ctt tgc gtg gag cga gtg ctg gat tgc agg acg gct ttt aag cct 5878
Asn Leu Cys Val Glu Arg Val Leu Asp Cys Arg Thr Ala Phe Lys Pro
1905 1910 1915 1920

gtg ctt gtg gat gaa ggg agg aag gtg gca ata aaa ggg cca ctt cgc 5926
Val Leu Val Asp Glu Gly Arg Lys Val Ala Ile Lys Gly Pro Leu Arg
1925 1930 1935

atc tcc gca tcc tct gct gct caa agg agg ggg cgc att ggg aga aat 5974
Ile Ser Ala Ser Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn
1940 1945 1950

ccc aac aga gat gga gac tca tac tac tat tct gag cct aca agt gaa 6022
Pro Asn Arg Asp Gly Asp Ser Tyr Tyr Tyr Ser Glu Pro Thr Ser Glu
1955 1960 1965

gat aat gcc cac cac gtc tgc tgg ttg gag gcc tca atg ctc ttg gac 6070
Asp Asn Ala His His Val Cys Trp Leu Glu Ala Ser Met Leu Leu Asp
1970 1975 1980

aac atg gag gtg agg ggt gga atg gtc gcc cca ctc tat ggc gtt gaa 6118
Asn Met Glu Val Arg Gly Gly Met Val Ala Pro Leu Tyr Gly Val Glu
1985 1990 1995 2000

gga act aaa aca cca gtt tcc cct ggt gaa atg aga ctg agg gat gac 6166
Gly Thr Lys Thr Pro Val Ser Pro Gly Glu Met Arg Leu Arg Asp Asp
2005 2010 2015

cag agg aaa gtc ttc aga gaa cta gtg agg aat tgt gac ctg ccc gtt 6214
Gln Arg Lys Val Phe Arg Glu Leu Val Arg Asn Cys Asp Leu Pro Val
2020 2025 2030

tgg ctt tcg tgg caa gtg gcc aag gct ggt ttg aag acg aat gat cgt 6262
Trp Leu Ser Trp Gln Val Ala Lys Ala Gly Leu Lys Thr Asn Asp Arg
2035 2040 2045

aag tgg tgt ttt gaa ggc cct gag gaa cat gag atc ttg aat gac agc 6310
Lys Trp Cys Phe Glu Gly Pro Glu Glu His Glu Ile Leu Asn Asp Ser
2050 2055 2060

ggt gaa aca gtg aag tgc agg gct cct gga gga gca aag aag cct ctg 6358
Gly Glu Thr Val Lys Cys Arg Ala Pro Gly Gly Ala Lys Lys Pro Leu
2065 2070 2075 2080

cgc cca agg tgg tgt gat gaa agg gtg tca tct gac cag agt gcg ctg 6406
Arg Pro Arg Trp Cys Asp Glu Arg Val Ser Ser Asp Gln Ser Ala Leu
2085 2090 2095

tct gaa ttt att aag ttt gct gaa ggt agg agg gga gct gcg gaa gtg 6454
Ser Glu Phe Ile Lys Phe Ala Glu Gly Arg Arg Gly Ala Ala Glu Val
2100 2105 2110

cta gtt gtg ctg agt gaa ctc cct gat ttc ctg gct aaa aaa ggt gga 6502
Leu Val Val Leu Ser Glu Leu Pro Asp Phe Leu Ala Lys Lys Gly Gly
2115 2120 2125

gag gca atg gat acc atc agt gtg ttt ctc cac tct gag gaa ggc tct 6550
Glu Ala Met Asp Thr Ile Ser Val Phe Leu His Ser Glu Glu Gly Ser
2130 2135 2140

agg gct tac cgc aat gca cta tca atg atg cct gag gca atg aca ata 6598
Arg Ala Tyr Arg Asn Ala Leu Ser Met Met Pro Glu Ala Met Thr Ile
2145 2150 2155 2160

gtc atg ctg ttt ata ctg gct gga cta ctg aca tcg gga atg gtc atc 6646
Val Met Leu Phe Ile Leu Ala Gly Leu Leu Thr Ser Gly Met Val Ile
2165 2170 2175

ttt ttc atg tct ccc aaa ggc atc agt aga atg tct atg gcg atg ggc 6694
Phe Phe Met Ser Pro Lys Gly Ile Ser Arg Met Ser Met Ala Met Gly
2180 2185 2190

aca atg gcc ggc tgt gga tat ctc atg ttc ctt gga ggc gtc aaa ccc 6742
Thr Met Ala Gly Cys Gly Tyr Leu Met Phe Leu Gly Gly Val Lys Pro
2195 2200 2205

act cac atc tcc tat atc atg ctc ata ttc ttt gtc ctg atg gtg gtt 6790
Thr His Ile Ser Tyr Ile Met Leu Ile Phe Phe Val Leu Met Val Val
2210 2215 2220

gtg atc ccc gag cca ggg caa caa agg tcc atc caa gac aac caa gtg 6838
Val Ile Pro Glu Pro Gly Gln Gln Arg Ser Ile Gln Asp Asn Gln Val
2225 2230 2235 2240

gca tac ctc att att ggc atc ctg acg ctg gtt tca gtg gtg gca gcc 6886
Ala Tyr Leu Ile Ile Gly Ile Leu Thr Leu Val Ser Val Val Ala Ala
2245 2250 2255

aac gag cta ggc atg ctg gag aaa acc aaa gag gac ctc ttt ggg aag 6934
Asn Glu Leu Gly Met Leu Glu Lys Thr Lys Glu Asp Leu Phe Gly Lys
2260 2265 2270

aag aac tta att cca tct agt gct tca ccc tgg agt tgg ccg gat ctt 6982
Lys Asn Leu Ile Pro Ser Ser Ala Ser Pro Trp Ser Trp Pro Asp Leu
2275 2280 2285

gac ctg aag cca gga gct gcc tgg aca gtg tac gtt ggc att gtt aca 7030
Asp Leu Lys Pro Gly Ala Ala Trp Thr Val Tyr Val Gly Ile Val Thr

2290 2295 2300

atg ctc tct cca atg ttg cac cac tgg atc aaa gtc gaa tat ggc aac 7078
Met Leu Ser Pro Met Leu His His Trp Ile Lys Val Glu Tyr Gly Asn
2305 2310 2315 2320

ctg tct ctg tct gga ata gcc cag tca gcc tca gtc ctt tct ttc atg 7126
Leu Ser Leu Ser Gly Ile Ala Gln Ser Ala Ser Val Leu Ser Phe Met
2325 2330 2335

gac aag ggg ata cca ttc atg aag atg aat atc tcg gtc ata ata ctg 7174
Asp Lys Gly Ile Pro Phe Met Lys Met Asn Ile Ser Val Ile Ile Leu
2340 2345 2350

ctg atc agt ggc tgg aat tca ata aca gtg atg cct ctg ctc tgt ggc 7222
Leu Ile Ser Gly Trp Asn Ser Ile Thr Val Met Pro Leu Leu Cys Gly
2355 2360 2365

ata ggg tgc gcc atg ctc cac tgg tct ctc att tta cct gga atc aaa 7270
Ile Gly Cys Ala Met Leu His Trp Ser Leu Ile Leu Pro Gly Ile Lys
2370 2375 2380

gcg cag cag tca aag ctt gca cag aga agg gtg ttc cat ggc gtt gcc 7318
Ala Gln Gln Ser Lys Leu Ala Gln Arg Arg Val Phe His Gly Val Ala
2385 2390 2395 2400

aag aac cct gtg gtt gat ggg aat cca aca gtt gac att gag gaa gct 7366
Lys Asn Pro Val Val Asp Gly Asn Pro Thr Val Asp Ile Glu Glu Ala
2405 2410 2415

cct gaa atg cct gcc ctt tat gag aag aaa ctg gct cta tat ctc ctt 7414
Pro Glu Met Pro Ala Leu Tyr Glu Lys Lys Leu Ala Leu Tyr Leu Leu
2420 2425 2430

ctt gct ctc agc cta gct tct gtt gcc atg tgc aga acg ccc ttt tca 7462
Leu Ala Leu Ser Leu Ala Ser Val Ala Met Cys Arg Thr Pro Phe Ser
2435 2440 2445

ttg gct gaa ggc att gtc cta gca tca gct gcc tta ggg ccg ctc ata 7510
Leu Ala Glu Gly Ile Val Leu Ala Ser Ala Ala Leu Gly Pro Leu Ile
2450 2455 2460

gag gga aac acc agc ctt ctt tgg aat gga ccc atg gct gtc tcc atg 7558
Glu Gly Asn Thr Ser Leu Leu Trp Asn Gly Pro Met Ala Val Ser Met
2465 2470 2475 2480

aca gga gtc atg cgg ggg aat tac tat gct ttt gtg gga gtc atg tac 7606
Thr Gly Val Met Arg Gly Asn Tyr Tyr Ala Phe Val Gly Val Met Tyr
2485 2490 2495

aat cta tgg aag atg aaa act gga cgc cgg ggg agt gcg aat gga aaa 7654
Asn Leu Trp Lys Met Lys Thr Gly Arg Arg Gly Ser Ala Asn Gly Lys
2500 2505 2510

act ttg ggt gaa gtc tgg aag agg gaa ctg aat ctg ttg gac aag caa 7702

Thr Leu Gly Glu Val Trp Lys Arg Glu Leu Asn Leu Leu Asp Lys Gln
2515 2520 2525

cag ttt gag ttg tat aaa agg acc gac att gtg gag gtg gat cgt gat 7750
Gln Phe Glu Leu Tyr Lys Arg Thr Asp Ile Val Glu Val Asp Arg Asp
2530 2535 2540

acg gca cgc agg cat ttg gcc gaa ggg aag gtg gac acc ggg gtg gcg 7798
Thr Ala Arg Arg His Leu Ala Glu Gly Lys Val Asp Thr Gly Val Ala
2545 2550 2555 2560

gtc tcc agg ggg acc gca aag tta agg tgg ttc cat gag cgt ggc tat 7846
Val Ser Arg Gly Thr Ala Lys Leu Arg Trp Phe His Glu Arg Gly Tyr
2565 2570 2575

gtc aag ctg gaa ggt agg gtg att gac ctg ggg tgt ggc cgc gga ggc 7894
Val Lys Leu Glu Gly Arg Val Ile Asp Leu Gly Cys Gly Arg Gly Gly
2580 2585 2590

tgg tgt tac tac gct gct gcg caa aag gaa gtg agt ggg gtc aaa gga 7942
Trp Cys Tyr Tyr Ala Ala Ala Gln Lys Glu Val Ser Gly Val Lys Gly
2595 2600 2605

ttc act ctt gga aga gac ggc cat gag aaa ccc atg aat gtg caa agt 7990
Phe Thr Leu Gly Arg Asp Gly His Glu Lys Pro Met Asn Val Gln Ser
2610 2615 2620

ctg gga tgg aac atc att acc ttc aag gac aaa act gat atc cac cgc 8038
Leu Gly Trp Asn Ile Ile Thr Phe Lys Asp Lys Thr Asp Ile His Arg
2625 2630 2635 2640

cta gaa cca gtg aaa tgt gac acc ctt ttg tgt gac att gga gag tca 8086
Leu Glu Pro Val Lys Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser
2645 2650 2655

tca tcg tca tcg gtc aca gag ggg gaa agg acc gtg aga gtt ctt gat 8134
Ser Ser Ser Ser Val Thr Glu Gly Glu Arg Thr Val Arg Val Leu Asp
2660 2665 2670

act gta gaa aaa tgg ctg gct tgt ggg gtt gac aac ttc tgt gtg aag 8182
Thr Val Glu Lys Trp Leu Ala Cys Gly Val Asp Asn Phe Cys Val Lys
2675 2680 2685

gtg tta gct cca tac atg cca gat gtt ctc gag aaa ctg gaa ttg ctc 8230
Val Leu Ala Pro Tyr Met Pro Asp Val Leu Glu Lys Leu Glu Leu Leu
2690 2695 2700

caa agg agg ttt ggc gga aca gtg atc agg aac cct ctc tcc agg aat 8278
Gln Arg Arg Phe Gly Gly Thr Val Ile Arg Asn Pro Leu Ser Arg Asn
2705 2710 2715 2720

tcc act cat gaa atg tac tac gtg tct gga gcc cgc agc aat gtc aca 8326
Ser Thr His Glu Met Tyr Tyr Val Ser Gly Ala Arg Ser Asn Val Thr
2725 2730 2735

ttt act gtg aac caa aca tcc cgc ctc ctg atg agg aga atg agg cgt 8374
Phe Thr Val Asn Gln Thr Ser Arg Leu Leu Met Arg Arg Met Arg Arg
2740 2745 2750

cca act gga aaa gtg acc ctg gag gct gac gtc atc ctc cca att ggg 8422
Pro Thr Gly Lys Val Thr Leu Glu Ala Asp Val Ile Leu Pro Ile Gly
2755 2760 2765

aca cgc agt gtt gag aca gac aag gga ccc ctg gac aaa gag gcc ata 8470
Thr Arg Ser Val Glu Thr Asp Lys Gly Pro Leu Asp Lys Glu Ala Ile
2770 2775 2780

gaa gaa agg gtt gag agg ata aaa tct gag tac atg acc tct tgg ttt 8518
Glu Glu Arg Val Glu Arg Ile Lys Ser Glu Tyr Met Thr Ser Trp Phe
2785 2790 2795 2800

tat gac aat gac aac ccc tac agg acc tgg cac tac tgt ggc tcc tat 8566
Tyr Asp Asn Asp Asn Pro Tyr Arg Thr Trp His Tyr Cys Gly Ser Tyr
2805 2810 2815

gtc aca aaa acc tca gga agt gcg gcg agc atg gta aat ggt gtt att 8614
Val Thr Lys Thr Ser Gly Ser Ala Ala Ser Met Val Asn Gly Val Ile
2820 2825 2830

aaa att ctg aca tac cca tgg gac agg ata gag gag gtc aca aga atg 8662
Lys Ile Leu Thr Tyr Pro Trp Asp Arg Ile Glu Glu Val Thr Arg Met
2835 2840 2845

gca atg act gac aca acc cct ttt gga cag caa aga gtg ttt aaa gaa 8710
Ala Met Thr Asp Thr Thr Pro Phe Gly Gln Gln Arg Val Phe Lys Glu
2850 2855 2860

aaa gtt gac acc aga gca aag gat cca cca gcg gga act agg aag atc 8758
Lys Val Asp Thr Arg Ala Lys Asp Pro Pro Ala Gly Thr Arg Lys Ile
2865 2870 2875 2880

atg aaa gtt gtc aac agg tgg ctg ttc cgc cac ctg gcc aga gaa aag 8806
Met Lys Val Val Asn Arg Trp Leu Phe Arg His Leu Ala Arg Glu Lys
2885 2890 2895

aac ccc aga ctg tgc aca aag gaa gaa ttt att gca aaa gtc cga agt 8854
Asn Pro Arg Leu Cys Thr Lys Glu Glu Phe Ile Ala Lys Val Arg Ser
2900 2905 2910

cat gca gcc att gga gct tac ctg gaa gaa caa gaa cag tgg aag act 8902
His Ala Ala Ile Gly Ala Tyr Leu Glu Glu Gln Glu Gln Trp Lys Thr
2915 2920 2925

gcc aat gag gct gtt caa gac cca aag ttc tgg gaa ctg gtg gat gaa 8950
Ala Asn Glu Ala Val Gln Asp Pro Lys Phe Trp Glu Leu Val Asp Glu
2930 2935 2940

gaa agg aag ctg cac caa caa ggc agg tgt cgg act tgt gtg tac aac 8998
Glu Arg Lys Leu His Gln Gln Gly Arg Cys Arg Thr Cys Val Tyr Asn
2945 2950 2955 2960

atg atg ggg aaa aga gag aag aag ctg tca gag ttt ggg aaa gca aag 9046
Met Met Gly Lys Arg Glu Lys Lys Leu Ser Glu Phe Gly Lys Ala Lys
2965 2970 2975

gga agc cgt gcc ata tgg tat atg tgg ctg gga gcg cgg tat ctt gag 9094
Gly Ser Arg Ala Ile Trp Tyr Met Trp Leu Gly Ala Arg Tyr Leu Glu
2980 2985 2990

ttt gag gcc ctg gga ttc ctg aat gag gac cat tgg gct tcc agg gaa 9142
Phe Glu Ala Leu Gly Phe Leu Asn Glu Asp His Trp Ala Ser Arg Glu
2995 3000 3005

aac tca gga gga gga gtg gaa ggc att ggc tta caa tac cta gga tat 9190
Asn Ser Gly Gly Gly Val Glu Gly Ile Gly Leu Gln Tyr Leu Gly Tyr
3010 3015 3020

gtg atc aga gac ctg gct gca atg gat ggt ggt gga ttc tac gcg gat 9238
Val Ile Arg Asp Leu Ala Ala Met Asp Gly Gly Gly Phe Tyr Ala Asp
3025 3030 3035 3040

gac acc gct gga tgg gac acg cgc atc aca gag gca gac ctt gat gat 9286
Asp Thr Ala Gly Trp Asp Thr Arg Ile Thr Glu Ala Asp Leu Asp Asp
3045 3050 3055

gaa cag gag atc ttg aac tac atg agc cca cat cac aaa aaa ctg gca 9334
Glu Gln Glu Ile Leu Asn Tyr Met Ser Pro His His Lys Lys Leu Ala
3060 3065 3070

caa gca gtg atg gaa atg aca tac aag aac aaa gtg gtg aaa gtg ttg 9382
Gln Ala Val Met Glu Met Thr Tyr Lys Asn Lys Val Val Lys Val Leu
3075 3080 3085

aga cca gcc cca gga ggg aaa gcc tac atg gat gtc ata agt cga cga 9430
Arg Pro Ala Pro Gly Gly Lys Ala Tyr Met Asp Val Ile Ser Arg Arg
3090 3095 3100

gac cag aga gga tcc ggg cag gta gtg act tat gct ctg aac acc atc 9478
Asp Gln Arg Gly Ser Gly Gln Val Val Thr Tyr Ala Leu Asn Thr Ile
3105 3110 3115 3120

acc aac ttg aaa gtc caa ttg atc aga atg gca gaa gca gag atg gtg 9526
Thr Asn Leu Lys Val Gln Leu Ile Arg Met Ala Glu Ala Glu Met Val
3125 3130 3135

ata cat cac caa cat gtt caa gat tgt gat gaa tca gtt ctg acc agg 9574
Ile His His Gln His Val Gln Asp Cys Asp Glu Ser Val Leu Thr Arg
3140 3145 3150

ctg gag gca tgg ctc act gag cac gga tgt aac aga ctg aag agg atg 9622
Leu Glu Ala Trp Leu Thr Glu His Gly Cys Asn Arg Leu Lys Arg Met
3155 3160 3165

gcg gtg agt gga gac gac tgt gtg gtc cgg ccc atc gat gac agg ttc 9670
Ala Val Ser Gly Asp Asp Cys Val Val Arg Pro Ile Asp Asp Arg Phe

3170 3175 3180

ggc ctg gcc ctg tcc cat ctc aac gcc atg tcc aag gtt aga aag gac 9718
Gly Leu Ala Leu Ser His Leu Asn Ala Met Ser Lys Val Arg Lys Asp
3185 3190 3195 3200

ata tct gaa tgg cag cca tca aaa ggg tgg aat gat tgg gag aat gtg 9766
Ile Ser Glu Trp Gln Pro Ser Lys Gly Trp Asn Asp Trp Glu Asn Val
3205 3210 3215

ccc ttc tgt tcc cac cac ttc cat gaa cta cag ctg aag gat ggc agg 9814
Pro Phe Cys Ser His His Phe His Glu Leu Gln Leu Lys Asp Gly Arg
3220 3225 3230

agg att gtg gtg cct tgc cga gaa cag gac gag ctc att ggg aga gga 9862
Arg Ile Val Val Pro Cys Arg Glu Gln Asp Glu Leu Ile Gly Arg Gly
3235 3240 3245

agg gtg tct cca gga aac ggc tgg atg atc aag gaa aca gct tgc ctc 9910
Arg Val Ser Pro Gly Asn Gly Trp Met Ile Lys Glu Thr Ala Cys Leu
3250 3255 3260

agc aaa gcc tat gcc aac atg tgg tca ctg atg tat ttt cac aaa agg 9958
Ser Lys Ala Tyr Ala Asn Met Trp Ser Leu Met Tyr Phe His Lys Arg
3265 3270 3275 3280

gac atg agg cta ctg tca ttg gct gtt tcc tca gct gtt ccc acc tca 10006
Asp Met Arg Leu Leu Ser Leu Ala Val Ser Ser Ala Val Pro Thr Ser
3285 3290 3295

tgg gtt cca caa gga cgc aca aca tgg tcg att cat ggg aaa ggg gag 10054
Trp Val Pro Gln Gly Arg Thr Thr Trp Ser Ile His Gly Lys Gly Glu
3300 3305 3310

tgg atg acc acg gaa gac atg ctt gag gtg tgg aac aga gta tgg ata 10102
Trp Met Thr Thr Glu Asp Met Leu Glu Val Trp Asn Arg Val Trp Ile
3315 3320 3325

acc aac aac cca cac atg cag gac aag aca atg gtg aaa gaa tgg aga 10150
Thr Asn Asn Pro His Met Gln Asp Lys Thr Met Val Lys Glu Trp Arg
3330 3335 3340

gat gtc cct tat cta acc aag aga caa gac aag ctg tgc gga tca ctg 10198
Asp Val Pro Tyr Leu Thr Lys Arg Gln Asp Lys Leu Cys Gly Ser Leu
3345 3350 3355 3360

att gga atg acc aat agg gcc acc tgg gcc tcc cac atc cat ttg gtc 10246
Ile Gly Met Thr Asn Arg Ala Thr Trp Ala Ser His Ile His Leu Val
3365 3370 3375

atc cat cgt atc cga acg ctg att gga cag gag aaa tat act gac tac 10294
Ile His Arg Ile Arg Thr Leu Ile Gly Gln Glu Lys Tyr Thr Asp Tyr
3380 3385 3390

cta aca gtc atg gac aga tat tct gtg gat gct gac ctg caa ccg ggt 10342

Leu Thr Val Met Asp Arg Tyr Ser Val Asp Ala Asp Leu Gln Pro Gly
3395 3400 3405

gag ctt atc tga aac acc atc taa tag gaa taa ccg gga tac aaa cca 10390
Glu Leu Ile Asn Thr Ile Glu Pro Gly Tyr Lys Pro
3410 3415 3420

cgg gtg gag aac cgg act ccc cac aac ttg aaa ccg gga tat aaa cca 10438
Arg Val Glu Asn Arg Thr Pro His Asn Leu Lys Pro Gly Tyr Lys Pro
3425 3430 3435 3440

cgg ctg gag aac cgg act ccg cac tta aaa tga aac aga aac cgg gat 10486
Arg Leu Glu Asn Arg Thr Pro His Leu Lys Asn Arg Asn Arg Asp
3445 3450 3455

aaa aac tac gga tgg aga acc gga ctc cac aca ttg aga cag aag aag 10534
Lys Asn Tyr Gly Trp Arg Thr Gly Leu His Thr Leu Arg Gln Lys Lys
3460 3465 3470

ttg tca gcc cag aac tcc aca cga gtt ttg cca ctg cta agc tgt gag 10582
Leu Ser Ala Gln Asn Ser Thr Arg Val Leu Pro Leu Leu Ser Cys Glu
3475 3480 3485

gca gtg cag gct ggg aca gcc gac ctc cag gtt gcg aaa aac ctg gtt 10630
Ala Val Gln Ala Gly Thr Ala Asp Leu Gln Val Ala Lys Asn Leu Val
3490 3495 3500

tct ggg acc tcc cac ccc aga gta aaa aga acg gag cct ccg cta cca 10678
Ser Gly Thr Ser His Pro Arg Val Lys Arg Thr Glu Pro Pro Leu Pro
3505 3510 3515 3520

ccc tcc cac gtg gtg gta gaa aga cgg ggt cta gag gtt aga gga gac 10726
Pro Ser His Val Val Val Glu Arg Arg Gly Leu Glu Val Arg Gly Asp
3525 3530 3535

cct cca ggg aac aaa tag tgggaccata ttgacgccag ggaaagaccg 10774
Pro Pro Gly Asn Lys
3540

gagtggttct ctgcttttcc tccaggggtc tgtgagcaca gtttgctcaa gaataagcag 10834

acctttggat gaaaaacaca aaaccact 10862

<210> 2
<211> 3411
<212> PRT
<213> Yellow fever virus

<400> 2
Met Ser Gly Arg Lys Ala Gln Gly Lys Thr Leu Gly Val Asn Met Val
1 5 10 15
Arg Arg Gly Val Arg Ser Leu Ser Asn Lys Ile Lys Gln Lys Thr Lys
20 25 30
Gln Ile Gly Asn Arg Pro Gly Pro Ser Arg Gly Val Gln Gly Phe Ile

35 40 45
Phe Phe Phe Leu Phe Asn Ile Leu Thr Gly Lys Lys Ile Thr Ala His
50 55 60
Leu Lys Arg Leu Trp Lys Met Leu Asp Pro Arg Gln Gly Leu Ala Val
65 70 75 80
Leu Arg Lys Val Lys Arg Val Val Ala Ser Leu Met Arg Gly Leu Ser
85 90 95
Ser Arg Lys Arg Arg Ser His Asp Val Leu Thr Val Gln Phe Leu Ile
100 105 110
Leu Gly Met Leu Leu Met Thr Gly Gly Val Thr Leu Val Arg Lys Asn
115 120 125
Arg Trp Leu Leu Leu Asn Val Thr Ser Glu Asp Leu Gly Lys Thr Phe
130 135 140
Ser Val Gly Thr Gly Asn Cys Thr Thr Asn Ile Leu Glu Ala Lys Tyr
145 150 155 160
Trp Cys Pro Asp Ser Met Glu Tyr Asn Cys Pro Asn Leu Ser Pro Arg
165 170 175
Glu Glu Pro Asp Asp Ile Asp Cys Trp Cys Tyr Gly Val Glu Asn Val
180 185 190
Arg Val Ala Tyr Gly Lys Cys Asp Ser Ala Gly Arg Ser Arg Arg Ser
195 200 205
Arg Arg Ala Ile Asp Leu Pro Thr His Glu Asn His Gly Leu Lys Thr
210 215 220
Arg Gln Glu Lys Trp Met Thr Gly Arg Met Gly Glu Arg Gln Leu Gln
225 230 235 240
Lys Ile Glu Arg Trp Leu Val Arg Asn Pro Phe Phe Ala Val Thr Ala
245 250 255
Leu Thr Ile Ala Tyr Leu Val Gly Ser Asn Met Thr Gln Arg Val Val
260 265 270
Ile Ala Leu Leu Val Leu Ala Val Gly Pro Ala Tyr Ser Ala His Cys
275 280 285
Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His Gly Gly Thr
290 295 300
Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr Val Met Ala
305 310 315 320
Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val Ala Ile Asp
325 330 335
Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val Leu Thr His
340 345 350
Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala His Leu Ala
355 360 365
Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr Ser Asp Arg
370 375 380
Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser Ile Val Ala
385 390 395 400
Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe Glu Val Asp
405 410 415
Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His Val Gly Ala
420 425 430
Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys Phe Asp Ala
435 440 445
Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly Lys Ala Thr
450 455 460
Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn Ser Tyr Ile
465 470 475 480

Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln Trp Ala Gln
485 490 495

Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val Trp Arg Glu
500 505 510

Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala Thr Ile Arg
515 520 525

Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr Ala Leu Thr
530 535 540

Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn Leu Tyr Lys
545 550 555 560

Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser Ala Leu Thr
565 570 575

Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met Ser Phe Val
580 585 590

Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met Gln Val Arg
595 600 605

Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val Ala Asp Asp
610 615 620

Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val Asn Pro Ile
625 630 635 640

Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn Pro Pro Phe
645 650 655

Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg Leu Thr Tyr
660 665 670

Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe Thr Gln Thr
675 680 685

Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala Ala Trp Asp
690 695 700

Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys Gly Ile His
705 710 715 720

Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly Leu Asn Trp
725 730 735

Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val Gly Ile Asn
740 745 750

Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val Gly Val Ile
755 760 765

Met Met Phe Leu Ser Leu Gly Val Gly Ala Asp Gln Gly Cys Ala Ile
770 775 780

Asn Phe Gly Lys Arg Glu Leu Lys Cys Gly Asp Gly Ile Phe Ile Phe
785 790 795 800

Arg Asp Ser Asp Asp Trp Leu Asn Lys Tyr Ser Tyr Tyr Pro Glu Asp
805 810 815

Pro Val Lys Leu Ala Ser Ile Val Lys Ala Ser Phe Glu Glu Gly Lys
820 825 830

Cys Gly Leu Asn Ser Val Asp Ser Leu Glu His Glu Met Trp Arg Ser
835 840 845

Arg Ala Asp Glu Ile Asn Ala Ile Leu Glu Glu Asn Glu Val Asp Ile
850 855 860

Ser Val Val Val Gln Asp Pro Lys Asn Val Tyr Gln Arg Gly Thr His
865 870 875 880

Pro Phe Ser Arg Ile Arg Asp Gly Leu Gln Tyr Gly Trp Lys Thr Trp
885 890 895

Gly Lys Asn Leu Val Phe Ser Pro Gly Arg Lys Asn Gly Ser Phe Ile
900 905 910

Ile Asp Gly Lys Ser Arg Lys Glu Cys Pro Phe Ser Asn Arg Val Trp

915 920 925
Asn Ser Phe Gln Ile Glu Glu Phe Gly Thr Gly Val Phe Thr Thr Arg
930 935 940
Val Tyr Met Asp Ala Val Phe Glu Tyr Thr Ile Asp Cys Asp Gly Ser
945 950 955 960
Ile Leu Gly Ala Ala Val Asn Gly Lys Lys Ser Ala His Gly Ser Pro
965 970 975
Thr Phe Trp Met Gly Ser His Glu Val Asn Gly Thr Trp Met Ile His
980 985 990
Thr Leu Glu Ala Leu Asp Tyr Lys Glu Cys Glu Trp Pro Leu Thr His
995 1000 1005
Thr Ile Gly Thr Ser Val Glu Glu Ser Glu Met Phe Met Pro Arg Ser
1010 1015 1020
Ile Gly Gly Pro Val Ser Ser His Asn His Ile Pro Gly Tyr Lys Val
1025 1030 1035 1040
Gln Thr Asn Gly Pro Trp Met Gln Val Pro Leu Glu Val Lys Arg Glu
1045 1050 1055
Ala Cys Pro Gly Thr Ser Val Ile Ile Asp Gly Asn Cys Asp Gly Arg
1060 1065 1070
Gly Lys Ser Thr Arg Ser Thr Thr Asp Ser Gly Lys Ile Ile Pro Glu
1075 1080 1085
Trp Cys Cys Arg Ser Cys Thr Met Pro Pro Val Ser Phe His Gly Ser
1090 1095 1100
Asp Gly Cys Trp Tyr Pro Met Glu Ile Arg Pro Arg Lys Thr His Glu
1105 1110 1115 1120
Ser His Leu Val Arg Ser Trp Val Thr Ala Gly Glu Ile His Ala Val
1125 1130 1135
Pro Phe Gly Leu Val Ser Met Met Ile Ala Met Glu Val Val Leu Arg
1140 1145 1150
Lys Arg Gln Gly Pro Lys Gln Met Leu Val Gly Gly Val Val Leu Leu
1155 1160 1165
Gly Ala Met Leu Val Gly Gln Val Thr Leu Leu Asp Leu Leu Lys Leu
1170 1175 1180
Thr Val Ala Val Gly Leu His Phe His Glu Met Asn Asn Gly Gly Asp
1185 1190 1195 1200
Ala Met Tyr Met Ala Leu Ile Ala Ala Phe Ser Ile Arg Pro Gly Leu
1205 1210 1215
Leu Ile Gly Phe Gly Leu Arg Thr Leu Trp Ser Pro Arg Glu Arg Leu
1220 1225 1230
Val Leu Ala Leu Gly Ala Ala Met Val Glu Ile Ala Leu Gly Gly Met
1235 1240 1245
Met Gly Gly Leu Trp Lys Tyr Leu Asn Ala Val Ser Leu Cys Ile Leu
1250 1255 1260
Thr Ile Asn Ala Val Ala Ser Arg Lys Ala Ser Asn Thr Ile Leu Pro
1265 1270 1275 1280
Leu Met Ala Leu Leu Thr Pro Val Thr Met Ala Glu Val Arg Leu Ala
1285 1290 1295
Thr Met Leu Phe Cys Thr Val Val Ile Ile Gly Val Leu His Gln Asn
1300 1305 1310
Ser Lys Asp Thr Ser Met Gln Lys Thr Ile Pro Leu Val Ala Leu Thr
1315 1320 1325
Leu Thr Ser Tyr Leu Gly Leu Thr Gln Pro Phe Leu Gly Leu Cys Ala
1330 1335 1340
Phe Leu Ala Thr Arg Ile Phe Gly Arg Arg Ser Ile Pro Val Asn Glu
1345 1350 1355 1360

Ala Leu Ala Ala Ala Gly Leu Val Gly Val Leu Ala Gly Leu Ala Phe
1365 1370 1375
Gln Glu Met Glu Asn Phe Leu Gly Pro Ile Ala Val Gly Gly Ile Leu
1380 1385 1390
Met Met Leu Val Ser Val Ala Gly Arg Val Asp Gly Leu Glu Leu Lys
1395 1400 1405
Lys Leu Gly Glu Val Ser Trp Glu Glu Glu Ala Glu Ile Ser Gly Ser
1410 1415 1420
Ser Ala Arg Tyr Asp Val Ala Leu Ser Glu Gln Gly Glu Phe Lys Leu
1425 1430 1435 1440
Leu Ser Glu Glu Lys Val Pro Trp Asp Gln Val Val Met Thr Ser Leu
1445 1450 1455
Ala Leu Val Gly Ala Ala Ile His Pro Phe Ala Leu Leu Leu Val Leu
1460 1465 1470
Ala Gly Trp Leu Phe His Val Arg Gly Ala Arg Arg Ser Gly Asp Val
1475 1480 1485
Leu Trp Asp Ile Pro Thr Pro Lys Ile Ile Glu Glu Cys Glu His Leu
1490 1495 1500
Glu Asp Gly Ile Tyr Gly Ile Phe Gln Ser Thr Phe Leu Gly Ala Ser
1505 1510 1515 1520
Gln Arg Gly Val Gly Val Ala Gln Gly Gly Val Phe His Thr Met Trp
1525 1530 1535
His Val Thr Arg Gly Ala Phe Leu Val Arg Asn Gly Lys Lys Leu Ile
1540 1545 1550
Pro Ser Trp Ala Ser Val Lys Glu Asp Leu Val Ala Tyr Gly Gly Ser
1555 1560 1565
Trp Lys Leu Glu Gly Arg Trp Asp Gly Glu Glu Glu Val Gln Leu Ile
1570 1575 1580
Ala Ala Val Pro Gly Lys Asn Val Val Asn Val Gln Thr Lys Pro Ser
1585 1590 1595 1600
Leu Phe Lys Val Arg Asn Gly Gly Glu Ile Gly Ala Val Ala Leu Asp
1605 1610 1615
Tyr Pro Ser Gly Thr Ser Gly Ser Pro Ile Val Asn Arg Asn Gly Glu
1620 1625 1630
Val Ile Gly Leu Tyr Gly Asn Gly Ile Leu Val Gly Asp Asn Ser Phe
1635 1640 1645
Val Ser Ala Ile Ser Gln Thr Glu Val Lys Glu Glu Gly Lys Glu Glu
1650 1655 1660
Leu Gln Glu Ile Pro Thr Met Leu Lys Lys Gly Met Thr Thr Ile Leu
1665 1670 1675 1680
Asp Phe His Pro Gly Ala Gly Lys Thr Arg Arg Phe Leu Pro Gln Ile
1685 1690 1695
Leu Ala Glu Cys Ala Arg Arg Arg Leu Arg Thr Leu Val Leu Ala Pro
1700 1705 1710
Thr Arg Val Val Leu Ser Glu Met Lys Glu Ala Phe His Gly Leu Asp
1715 1720 1725
Val Lys Phe His Thr Gln Ala Phe Ser Ala His Gly Ser Gly Arg Glu
1730 1735 1740
Val Ile Asp Ala Met Cys His Ala Thr Leu Thr Tyr Arg Met Leu Glu
1745 1750 1755 1760
Pro Thr Arg Val Val Asn Trp Glu Val Ile Ile Met Asp Glu Ala His
1765 1770 1775
Phe Leu Asp Pro Ala Ser Ile Ala Ala Arg Gly Trp Ala Ala His Arg
1780 1785 1790
Ala Arg Ala Asn Glu Ser Ala Thr Ile Leu Met Thr Ala Thr Pro Pro

1795 1800 1805
Gly Thr Ser Asp Glu Phe Pro His Ser Asn Gly Glu Ile Glu Asp Val
1810 1815 1820
Gln Thr Asp Ile Pro Ser Glu Pro Trp Asn Thr Gly His Asp Trp Ile
1825 1830 1835 1840
Leu Ala Asp Lys Arg Pro Thr Ala Trp Phe Leu Pro Ser Ile Arg Ala
1845 1850 1855
Ala Asn Val Met Ala Ala Ser Leu Arg Lys Ala Gly Lys Ser Val Val
1860 1865 1870
Val Leu Asn Arg Lys Thr Phe Glu Arg Glu Tyr Pro Thr Ile Lys Gln
1875 1880 1885
Lys Lys Pro Asp Phe Ile Leu Ala Thr Asp Ile Ala Glu Met Gly Ala
1890 1895 1900
Asn Leu Cys Val Glu Arg Val Leu Asp Cys Arg Thr Ala Phe Lys Pro
1905 1910 1915 1920
Val Leu Val Asp Glu Gly Arg Lys Val Ala Ile Lys Gly Pro Leu Arg
1925 1930 1935
Ile Ser Ala Ser Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn
1940 1945 1950
Pro Asn Arg Asp Gly Asp Ser Tyr Tyr Tyr Ser Glu Pro Thr Ser Glu
1955 1960 1965
Asp Asn Ala His His Val Cys Trp Leu Glu Ala Ser Met Leu Leu Asp
1970 1975 1980
Asn Met Glu Val Arg Gly Gly Met Val Ala Pro Leu Tyr Gly Val Glu
1985 1990 1995 2000
Gly Thr Lys Thr Pro Val Ser Pro Gly Glu Met Arg Leu Arg Asp Asp
2005 2010 2015
Gln Arg Lys Val Phe Arg Glu Leu Val Arg Asn Cys Asp Leu Pro Val
2020 2025 2030
Trp Leu Ser Trp Gln Val Ala Lys Ala Gly Leu Lys Thr Asn Asp Arg
2035 2040 2045
Lys Trp Cys Phe Glu Gly Pro Glu Glu His Glu Ile Leu Asn Asp Ser
2050 2055 2060
Gly Glu Thr Val Lys Cys Arg Ala Pro Gly Gly Ala Lys Lys Pro Leu
2065 2070 2075 2080
Arg Pro Arg Trp Cys Asp Glu Arg Val Ser Ser Asp Gln Ser Ala Leu
2085 2090 2095
Ser Glu Phe Ile Lys Phe Ala Glu Gly Arg Arg Gly Ala Ala Glu Val
2100 2105 2110
Leu Val Val Leu Ser Glu Leu Pro Asp Phe Leu Ala Lys Lys Gly Gly
2115 2120 2125
Glu Ala Met Asp Thr Ile Ser Val Phe Leu His Ser Glu Glu Gly Ser
2130 2135 2140
Arg Ala Tyr Arg Asn Ala Leu Ser Met Met Pro Glu Ala Met Thr Ile
2145 2150 2155 2160
Val Met Leu Phe Ile Leu Ala Gly Leu Leu Thr Ser Gly Met Val Ile
2165 2170 2175
Phe Phe Met Ser Pro Lys Gly Ile Ser Arg Met Ser Met Ala Met Gly
2180 2185 2190
Thr Met Ala Gly Cys Gly Tyr Leu Met Phe Leu Gly Gly Val Lys Pro
2195 2200 2205
Thr His Ile Ser Tyr Ile Met Leu Ile Phe Phe Val Leu Met Val Val
2210 2215 2220
Val Ile Pro Glu Pro Gly Gln Gln Arg Ser Ile Gln Asp Asn Gln Val
2225 2230 2235 2240

Ala Tyr Leu Ile Ile Gly Ile Leu Thr Leu Val Ser Val Val Ala Ala
2245 2250 2255
Asn Glu Leu Gly Met Leu Glu Lys Thr Lys Glu Asp Leu Phe Gly Lys
2260 2265 2270
Lys Asn Leu Ile Pro Ser Ser Ala Ser Pro Trp Ser Trp Pro Asp Leu
2275 2280 2285
Asp Leu Lys Pro Gly Ala Ala Trp Thr Val Tyr Val Gly Ile Val Thr
2290 2295 2300
Met Leu Ser Pro Met Leu His His Trp Ile Lys Val Glu Tyr Gly Asn
2305 2310 2315 2320
Leu Ser Leu Ser Gly Ile Ala Gln Ser Ala Ser Val Leu Ser Phe Met
2325 2330 2335
Asp Lys Gly Ile Pro Phe Met Lys Met Asn Ile Ser Val Ile Ile Leu
2340 2345 2350
Leu Ile Ser Gly Trp Asn Ser Ile Thr Val Met Pro Leu Leu Cys Gly
2355 2360 2365
Ile Gly Cys Ala Met Leu His Trp Ser Leu Ile Leu Pro Gly Ile Lys
2370 2375 2380
Ala Gln Gln Ser Lys Leu Ala Gln Arg Arg Val Phe His Gly Val Ala
2385 2390 2395 2400
Lys Asn Pro Val Val Asp Gly Asn Pro Thr Val Asp Ile Glu Glu Ala
2405 2410 2415
Pro Glu Met Pro Ala Leu Tyr Glu Lys Lys Leu Ala Leu Tyr Leu Leu
2420 2425 2430
Leu Ala Leu Ser Leu Ala Ser Val Ala Met Cys Arg Thr Pro Phe Ser
2435 2440 2445
Leu Ala Glu Gly Ile Val Leu Ala Ser Ala Ala Leu Gly Pro Leu Ile
2450 2455 2460
Glu Gly Asn Thr Ser Leu Leu Trp Asn Gly Pro Met Ala Val Ser Met
2465 2470 2475 2480
Thr Gly Val Met Arg Gly Asn Tyr Tyr Ala Phe Val Gly Val Met Tyr
2485 2490 2495
Asn Leu Trp Lys Met Lys Thr Gly Arg Arg Gly Ser Ala Asn Gly Lys
2500 2505 2510
Thr Leu Gly Glu Val Trp Lys Arg Glu Leu Asn Leu Leu Asp Lys Gln
2515 2520 2525
Gln Phe Glu Leu Tyr Lys Arg Thr Asp Ile Val Glu Val Asp Arg Asp
2530 2535 2540
Thr Ala Arg Arg His Leu Ala Glu Gly Lys Val Asp Thr Gly Val Ala
2545 2550 2555 2560
Val Ser Arg Gly Thr Ala Lys Leu Arg Trp Phe His Glu Arg Gly Tyr
2565 2570 2575
Val Lys Leu Glu Gly Arg Val Ile Asp Leu Gly Cys Gly Arg Gly Gly
2580 2585 2590
Trp Cys Tyr Tyr Ala Ala Ala Gln Lys Glu Val Ser Gly Val Lys Gly
2595 2600 2605
Phe Thr Leu Gly Arg Asp Gly His Glu Lys Pro Met Asn Val Gln Ser
2610 2615 2620
Leu Gly Trp Asn Ile Ile Thr Phe Lys Asp Lys Thr Asp Ile His Arg
2625 2630 2635 2640
Leu Glu Pro Val Lys Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser
2645 2650 2655
Ser Ser Ser Ser Val Thr Glu Gly Glu Arg Thr Val Arg Val Leu Asp
2660 2665 2670
Thr Val Glu Lys Trp Leu Ala Cys Gly Val Asp Asn Phe Cys Val Lys

2675 2680 2685
Val Leu Ala Pro Tyr Met Pro Asp Val Leu Glu Lys Leu Glu Leu Leu
2690 2695 2700
Gln Arg Arg Phe Gly Gly Thr Val Ile Arg Asn Pro Leu Ser Arg Asn
2705 2710 2715 2720
Ser Thr His Glu Met Tyr Tyr Val Ser Gly Ala Arg Ser Asn Val Thr
2725 2730 2735
Phe Thr Val Asn Gln Thr Ser Arg Leu Leu Met Arg Arg Met Arg Arg
2740 2745 2750
Pro Thr Gly Lys Val Thr Leu Glu Ala Asp Val Ile Leu Pro Ile Gly
2755 2760 2765
Thr Arg Ser Val Glu Thr Asp Lys Gly Pro Leu Asp Lys Glu Ala Ile
2770 2775 2780
Glu Glu Arg Val Glu Arg Ile Lys Ser Glu Tyr Met Thr Ser Trp Phe
2785 2790 2795 2800
Tyr Asp Asn Asp Asn Pro Tyr Arg Thr Trp His Tyr Cys Gly Ser Tyr
2805 2810 2815
Val Thr Lys Thr Ser Gly Ser Ala Ala Ser Met Val Asn Gly Val Ile
2820 2825 2830
Lys Ile Leu Thr Tyr Pro Trp Asp Arg Ile Glu Glu Val Thr Arg Met
2835 2840 2845
Ala Met Thr Asp Thr Thr Pro Phe Gly Gln Gln Arg Val Phe Lys Glu
2850 2855 2860
Lys Val Asp Thr Arg Ala Lys Asp Pro Pro Ala Gly Thr Arg Lys Ile
2865 2870 2875 2880
Met Lys Val Val Asn Arg Trp Leu Phe Arg His Leu Ala Arg Glu Lys
2885 2890 2895
Asn Pro Arg Leu Cys Thr Lys Glu Glu Phe Ile Ala Lys Val Arg Ser
2900 2905 2910
His Ala Ala Ile Gly Ala Tyr Leu Glu Glu Gln Glu Gln Trp Lys Thr
2915 2920 2925
Ala Asn Glu Ala Val Gln Asp Pro Lys Phe Trp Glu Leu Val Asp Glu
2930 2935 2940
Glu Arg Lys Leu His Gln Gln Gly Arg Cys Arg Thr Cys Val Tyr Asn
2945 2950 2955 2960
Met Met Gly Lys Arg Glu Lys Lys Leu Ser Glu Phe Gly Lys Ala Lys
2965 2970 2975
Gly Ser Arg Ala Ile Trp Tyr Met Trp Leu Gly Ala Arg Tyr Leu Glu
2980 2985 2990
Phe Glu Ala Leu Gly Phe Leu Asn Glu Asp His Trp Ala Ser Arg Glu
2995 3000 3005
Asn Ser Gly Gly Gly Val Glu Gly Ile Gly Leu Gln Tyr Leu Gly Tyr
3010 3015 3020
Val Ile Arg Asp Leu Ala Ala Met Asp Gly Gly Gly Phe Tyr Ala Asp
3025 3030 3035 3040
Asp Thr Ala Gly Trp Asp Thr Arg Ile Thr Glu Ala Asp Leu Asp Asp
3045 3050 3055
Glu Gln Glu Ile Leu Asn Tyr Met Ser Pro His His Lys Lys Leu Ala
3060 3065 3070
Gln Ala Val Met Glu Met Thr Tyr Lys Asn Lys Val Val Lys Val Leu
3075 3080 3085
Arg Pro Ala Pro Gly Gly Lys Ala Tyr Met Asp Val Ile Ser Arg Arg
3090 3095 3100
Asp Gln Arg Gly Ser Gly Gln Val Val Thr Tyr Ala Leu Asn Thr Ile
3105 3110 3115 3120

Thr Asn Leu Lys Val Gln Leu Ile Arg Met Ala Glu Ala Glu Met Val
3125 3130 3135
Ile His His Gln His Val Gln Asp Cys Asp Glu Ser Val Leu Thr Arg
3140 3145 3150
Leu Glu Ala Trp Leu Thr Glu His Gly Cys Asn Arg Leu Lys Arg Met
3155 3160 3165
Ala Val Ser Gly Asp Asp Cys Val Val Arg Pro Ile Asp Asp Arg Phe
3170 3175 3180
Gly Leu Ala Leu Ser His Leu Asn Ala Met Ser Lys Val Arg Lys Asp
3185 3190 3195 3200
Ile Ser Glu Trp Gln Pro Ser Lys Gly Trp Asn Asp Trp Glu Asn Val
3205 3210 3215
Pro Phe Cys Ser His His Phe His Glu Leu Gln Leu Lys Asp Gly Arg
3220 3225 3230
Arg Ile Val Val Pro Cys Arg Glu Gln Asp Glu Leu Ile Gly Arg Gly
3235 3240 3245
Arg Val Ser Pro Gly Asn Gly Trp Met Ile Lys Glu Thr Ala Cys Leu
3250 3255 3260
Ser Lys Ala Tyr Ala Asn Met Trp Ser Leu Met Tyr Phe His Lys Arg
3265 3270 3275 3280
Asp Met Arg Leu Leu Ser Leu Ala Val Ser Ser Ala Val Pro Thr Ser
3285 3290 3295
Trp Val Pro Gln Gly Arg Thr Thr Trp Ser Ile His Gly Lys Gly Glu
3300 3305 3310
Trp Met Thr Thr Glu Asp Met Leu Glu Val Trp Asn Arg Val Trp Ile
3315 3320 3325
Thr Asn Asn Pro His Met Gln Asp Lys Thr Met Val Lys Glu Trp Arg
3330 3335 3340
Asp Val Pro Tyr Leu Thr Lys Arg Gln Asp Lys Leu Cys Gly Ser Leu
3345 3350 3355 3360
Ile Gly Met Thr Asn Arg Ala Thr Trp Ala Ser His Ile His Leu Val
3365 3370 3375
Ile His Arg Ile Arg Thr Leu Ile Gly Gln Glu Lys Tyr Thr Asp Tyr
3380 3385 3390
Leu Thr Val Met Asp Arg Tyr Ser Val Asp Ala Asp Leu Gln Pro Gly
3395 3400 3405
Glu Leu Ile
3410

<210> 3
<211> 1479
<212> DNA
<213> Yellow fever virus

<220>
<221> CDS
<222> (1)..(1479)

<400> 3
gct cac tgc att gga att act gac aga gat ttc att gag ggg gtg cat 48
Ala His Cys Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His

```
 1               5                   10                  15

gga gga act tgg gtt tca gct acc ctg gag cac ggc aag tgt gtc act     96
Gly Gly Thr Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr
            20                  25                  30

gtt atg gcc cct gac aag cct tca ttg gac atc tca cta gag aca gta    144
Val Met Ala Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val
        35                  40                  45

gcc att gat gga cct gct gag gcg agg aaa gtg tgt tac aat gca gtt    192
Ala Ile Asp Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val
    50                  55                  60

ctc act cat gtg aag att aat gac aag tgc ccc agc act gga gag gcc    240
Leu Thr His Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala
65                  70                  75                  80

cac cta gct gaa gag aac gaa ggg gac aat gcg tgc aag cgc act tat    288
His Leu Ala Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr
                85                  90                  95

tct gat aga ggc tgg ggc aat ggc tgt ggc cta ttt ggg aaa ggg agc    336
Ser Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser
            100                 105                 110

att gtg gca tgc gcc aaa ttc act tgt gcc aaa tcc atg agt ttg ttt    384
Ile Val Ala Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe
        115                 120                 125

gag gtt gat cag acc aaa att cag tat gtc atc aga gca caa ttg cat    432
Glu Val Asp Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His
    130                 135                 140

gta ggg gcc aag cag gaa aat tgg aat acc gcc att aag act ctc aag    480
Val Gly Ala Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys
145                 150                 155                 160

ttt gat gcc ctg tca ggc tcc cag gaa gcc gag ttc act ggg tat gga    528
Phe Asp Ala Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly
                165                 170                 175

aaa gct aca ctg gaa tgc cag gtg caa act gcg gtg gac ttt ggt aac    576
Lys Ala Thr Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn
            180                 185                 190

agt tac atc gct gag atg gaa aaa gag agc tgg ata gtg gac aga cag    624
Ser Tyr Ile Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln
        195                 200                 205

tgg gcc cag gac ttg acc ctg cca tgg cag agt gga agt ggc ggg gtg    672
Trp Ala Gln Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val
    210                 215                 220

tgg aga gag atg cat cat ctt gtc gaa ttt gaa cct ccg cat gcc gcc    720
```

```
Trp Arg Glu Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala
225                 230                 235                 240

act atc aga gta ctg gcc ctg gga aac cag gaa ggc tcc ttg aaa aca     768
Thr Ile Arg Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr
                245                 250                 255

gct ctt acc ggc gca atg agg gtt aca aag gac aca aat gac aac aac     816
Ala Leu Thr Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn
            260                 265                 270

ctt tac aaa cta cat ggt gga cat gtt tcc tgc aga gtg aaa ttg tca     864
Leu Tyr Lys Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser
        275                 280                 285

gct ttg aca ctc aag ggg aca tcc tac aaa atg tgc act gac aaa atg     912
Ala Leu Thr Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met
    290                 295                 300

tct ttt gtc aag aac cca act gac act ggc cat ggc act gtt gtg atg     960
Ser Phe Val Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met
305                 310                 315                 320

cag gtg aga gtg cca aaa gga gcc ccc tgc agg att cca gtg ata gta     1008
Gln Val Arg Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val
                325                 330                 335

gct gat gat ctt aca gcg gca atc aat aaa ggc att ttg gtt aca gtt     1056
Ala Asp Asp Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val
            340                 345                 350

aac ccc atc gcc tca acc aat gat gat gaa gtg ctg att gag gtg aac     1104
Asn Pro Ile Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn
        355                 360                 365

cca cct ttt gga gac agc tac att atc gtt ggg aca gga gat tca cgt     1152
Pro Pro Phe Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg
    370                 375                 380

ctc act tac cag tgg cac aaa gag gga agc tca ata gga aag ttg ttc     1200
Leu Thr Tyr Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe
385                 390                 395                 400

act cag acc atg aaa ggc gcg gaa cgc ctg gcc gtc atg gga gac gcc     1248
Thr Gln Thr Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala
                405                 410                 415

gcc tgg gat ttc agc tcc gct gga ggg ttc ttc act tcg gtt ggg aaa     1296
Ala Trp Asp Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys
            420                 425                 430

gga att cat acg gtg ttt ggc tct gcc ttt cag ggg cta ttt ggc ggc     1344
Gly Ile His Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly
        435                 440                 445
```

ttg aac tgg ata aca aag gtc atc atg ggg gcg gta ctc ata tgg gtt 1392
Leu Asn Trp Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val
450 455 460

ggc atc aac aca aga aac atg aca atg tcc atg agc atg atc ttg gta 1440
Gly Ile Asn Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val
465 470 475 480

gga gtg atc atg atg ttt ttg tct cta gga gtt ggg gcg 1479
Gly Val Ile Met Met Phe Leu Ser Leu Gly Val Gly Ala
485 490

<210> 4
<211> 493
<212> PRT
<213> Yellow fever virus

<400> 4
Ala His Cys Ile Gly Ile Thr Asp Arg Asp Phe Ile Glu Gly Val His
1 5 10 15

Gly Gly Thr Trp Val Ser Ala Thr Leu Glu His Gly Lys Cys Val Thr
20 25 30

Val Met Ala Pro Asp Lys Pro Ser Leu Asp Ile Ser Leu Glu Thr Val
35 40 45

Ala Ile Asp Gly Pro Ala Glu Ala Arg Lys Val Cys Tyr Asn Ala Val
50 55 60

Leu Thr His Val Lys Ile Asn Asp Lys Cys Pro Ser Thr Gly Glu Ala
65 70 75 80

His Leu Ala Glu Glu Asn Glu Gly Asp Asn Ala Cys Lys Arg Thr Tyr
85 90 95

Ser Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser
100 105 110

Ile Val Ala Cys Ala Lys Phe Thr Cys Ala Lys Ser Met Ser Leu Phe
115 120 125

Glu Val Asp Gln Thr Lys Ile Gln Tyr Val Ile Arg Ala Gln Leu His
130 135 140

Val Gly Ala Lys Gln Glu Asn Trp Asn Thr Ala Ile Lys Thr Leu Lys
145 150 155 160

Phe Asp Ala Leu Ser Gly Ser Gln Glu Ala Glu Phe Thr Gly Tyr Gly
165 170 175

Lys Ala Thr Leu Glu Cys Gln Val Gln Thr Ala Val Asp Phe Gly Asn
180 185 190

```
Ser Tyr Ile Ala Glu Met Glu Lys Glu Ser Trp Ile Val Asp Arg Gln
        195                 200                 205

Trp Ala Gln Asp Leu Thr Leu Pro Trp Gln Ser Gly Ser Gly Gly Val
    210                 215                 220

Trp Arg Glu Met His His Leu Val Glu Phe Glu Pro Pro His Ala Ala
225                 230                 235                     240

Thr Ile Arg Val Leu Ala Leu Gly Asn Gln Glu Gly Ser Leu Lys Thr
                245                 250                 255

Ala Leu Thr Gly Ala Met Arg Val Thr Lys Asp Thr Asn Asp Asn Asn
            260                 265                 270

Leu Tyr Lys Leu His Gly Gly His Val Ser Cys Arg Val Lys Leu Ser
        275                 280                 285

Ala Leu Thr Leu Lys Gly Thr Ser Tyr Lys Met Cys Thr Asp Lys Met
    290                 295                 300

Ser Phe Val Lys Asn Pro Thr Asp Thr Gly His Gly Thr Val Val Met
305                 310                 315                     320

Gln Val Arg Val Pro Lys Gly Ala Pro Cys Arg Ile Pro Val Ile Val
                325                 330                 335

Ala Asp Asp Leu Thr Ala Ala Ile Asn Lys Gly Ile Leu Val Thr Val
            340                 345                 350

Asn Pro Ile Ala Ser Thr Asn Asp Asp Glu Val Leu Ile Glu Val Asn
        355                 360                 365

Pro Pro Phe Gly Asp Ser Tyr Ile Ile Val Gly Thr Gly Asp Ser Arg
    370                 375                 380

Leu Thr Tyr Gln Trp His Lys Glu Gly Ser Ser Ile Gly Lys Leu Phe
385                 390                 395                     400

Thr Gln Thr Met Lys Gly Ala Glu Arg Leu Ala Val Met Gly Asp Ala
                405                 410                 415

Ala Trp Asp Phe Ser Ser Ala Gly Gly Phe Phe Thr Ser Val Gly Lys
            420                 425                 430

Gly Ile His Thr Val Phe Gly Ser Ala Phe Gln Gly Leu Phe Gly Gly
        435                 440                 445

Leu Asn Trp Ile Thr Lys Val Ile Met Gly Ala Val Leu Ile Trp Val
    450                 455                 460

Gly Ile Asn Thr Arg Asn Met Thr Met Ser Met Ser Met Ile Leu Val
465                 470                 475                     480
Gly Val Ile Met Met Phe Leu Ser Leu Gly Val Gly Ala
                485                 490
```